Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

2004-06

Metrics for success : using metrics in exercises to assess the preparedness of the fire service in Homeland Security

Doherty, Vincent J.

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/1164

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

METRICS FOR SUCCESS:
USING METRICS IN EXERCISES TO ASSESS
THE PREPAREDNESS OF THE FIRE SERVICE
IN HOMELAND SECURITY

by

Vincent J. Doherty

June 2004

Thesis Advisor:	Paul Stockton
Second Reader:	Rudy Darken

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY	**2. REPORT DATE** June 2004	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: METRICS FOR SUCCESS: USING METRICS IN EXERCISES TO ASSESS THE PREPAREDNESS OF THE FIRE SERVICE IN HOMELAND SECURITY		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Vincent J. Doherty			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES: The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited		**12b. DISTRIBUTION CODE**	

13. ABSTRACT ***(maximum 200 words*** The Fire Service is at an important crossroads in its history. The increasing threat of terrorism, along with the need for the Fire Service to make a major contribution to the Homeland Security effort, compels firefighters to expand upon their traditional mission requirements. They must deepen and widen the skills they possess for responding to all hazard incidents, and must also cover the terroristic events and weapons of mass destruction incidents that are inevitable during this war on terrorism. This will only be possible if the Fire Service embraces major conceptual changes in the way that it trains and exercises its members. In many respects, that exercise and training system is excellent and serves as a model for other disciplines. In other ways, however, it tends to be flawed and burdensome, preventing the Fire Service from taking major strides forward in its training efforts – especially with respect to the use of metrics to objectively evaluate performance capabilities for Homeland Security-related operations, and the embedding of assessment techniques in a broader system to provide for improved performance.

This thesis examines the current approach taken by the Fire Service to training and exercising, especially in the Fire Department of New York (FDNY), and identifies gaps and problems in those existing systems. Then, the thesis examines relevant "best practices" from the United States Military and the private sector that might be applied to the shortfalls in Fire Service training and exercising. The remainder of the thesis proposes changes to the training and exercise system, designed to make a measurable, sustained impact on the capabilities of the fire service to accomplish Homeland Security and traditional mission requirements.

14. SUBJECT TERMS Fire Service, Performance Measurement, FDNY, Operational Exercises, Metrics, HSEEP.			**15. NUMBER OF PAGES** 90
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18-298-102

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution in unlimited

METRICS FOR SUCCESS
USING METRICS IN EXERCISES TO ASSESS
THE PREPAREDNESS OF THE FIRE SERVICE
IN HOMELAND SECURITY

Vincent J. Doherty
Captain, Fire Department of New York City
B.A., St. John's University, 1978

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF ARTS IN SECURITY STUDIES,
(HOMELAND SECURITY AND DEFENSE)

from the

NAVAL POSTGRADUATE SCHOOL
June 2004

Author: Vincent J. Doherty

Approved by: Paul Stockton
Thesis Advisor

Rudy Darken
Second Reader

James J. Wirtz
Chairman
Department of National Security Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Fire Service is at an important crossroads in its history. The increasing threat of terrorism, along with the need for the Fire Service to make a major contribution to the Homeland Security effort, compels firefighters to expand upon their traditional mission requirements. They must deepen and widen the skills they possess for responding to all hazard incidents, and must also cover the terroristic events and weapons of mass destruction incidents that are inevitable during this war on terrorism. This will only be possible if the Fire Service embraces major conceptual changes in the way that it trains and exercises its members. In many respects, that exercise and training system is excellent and serves as a model for other disciplines. In other ways, however, it tends to be flawed and burdensome, preventing the Fire Service from taking major strides forward in its training efforts – especially with respect to the use of metrics to objectively evaluate performance capabilities for Homeland Security-related operations, and the embedding of assessment techniques in a broader system to provide for improved performance.

This thesis examines the current approach taken by the Fire Service to training and exercising, especially in the Fire Department of New York (FDNY), and identifies gaps and problems in those existing systems.Then, the thesis examines relevant “best practices” from the United States Military and the private sector that might be applied to the shortfalls in Fire Service training and exercising. The remainder of the thesis proposes changes to the training and exercise system, designed to make a measurable, sustained impact on the capabilities of the fire service to accomplish Homeland Security and traditional mission requirements.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I: INTRODUCTION……..……. 1
A. Purpose……..1
B. Thesis Argument……... 2
C. Recommendations of this Thesis……...4
D. Scope of Research……6
E. Research Methodology and Thesis Structure…….…….. 7

II: THE CURRENT EXERCISE SYSTEM OF THE FIRE SERVICE:
EXCELLENCE AND OPPORTUNITIES FOR IMPROVRMENT …….. 9
A. Introduction ……..…….9
B. The Homeland Security Exercise and Evaluation Program (HSEEP)…….. 11
1. Overview of the HSEEP ……..11
C. Current FDNY Training and Exercise Model ……. 13
D. Challenges Faced by the FDNY and the Fire Service ……. 16
E. Emerging Best Practices …….. 18
F. Shortfalls …….. 20

III: REVIEW OF BEST PRACTICES IN THE U.S. MILITARY, THE PRIVATE
SECTOR, AND AS DESCRIBED IN ACADEMIC LITERATURE …… 23
A. Introduction ……. 23
B. The United States Army's Evaluation and Assessment Model ……... 23
1. Overview of the Model …….. 23
2. Assessment and Evaluation …… 24
3. After Action Review (AAR) …….. 26
4. The Army's Ten Principles of Training ……. 27
5. The US Army's Train the Force Model-Good Points, Bad Points……. 27
C. Best Practices from the Private Sector …… 29
1. Justification for Examining Industrial Models…… 29
2. Performance Measurement Models……29
a. Kirkpatrick – Four Step Model……. 30
b. Hamblin – Five Step Model……. 31
c. Organizational Elements Model (OEM) – Six Steps…….. 31
d. Indiana University and the Carousel of development:
Industrial Society Model (IS) …… 32
e. Return on Investment (ROI) & KPMT Models…… 33
f. CIRO (Context, Input, Reaction, Outcome) Model…….. 33
g. Alternate Evaluation Models…… 33
h. Models Using Different Measures: Evaluation Models…….. 34
3. Summary: Private Sector Concepts That Can Be Adapted to the Fire Service. 35

IV: DEVELOPMENT OF METRICS …… 37
A. Introduction …… 37
B. Capability, Preparedness and Readiness ……. 37
C. Measuring Performance …….. 38
1. Why measure?..... 38
2. Who do you measure?..... 39
a. The Individual Participants…….. 39
b. Groups and teams……. 40
c. Supervisors and Incident Commanders…… 40
3. What do you measure?..... 41
4. When do you measure?..... 43
5. How do you measure?..... 45
6. Where do you measure?..... 45
D. Things to Remember When Developing Metrics…….…….. 46

V: THE PROPOSED METRIC SYSTEM. ……..…….. 49
A. Introduction…….. 49
B. EDT and ECPE Functions……..... 49
C. Exercise Design Criteria ……. 50
D. Types of Exercises to Use When Designing…… 51
E. Measures to Use in the Exercise…….. 53
F. Assessing Preparedness: Drawing on Leadership Wisdom……..…… 55
G. Evaluation System for the Fire Service…….….. 56
H Steps to Follow when assessing the results of an Exercise……..….... 57
1. Evaluation Performance Measure Index (EPMI) …….. 58
2. Relative Preparedness Index (RPI) …… 59
3. Sample Calculations …….. 60

VI: CONCLUSIONS AND RECOMMENDATIONS …….……. 65
A. Introduction……....….. 65
B. Recommendations……..… 65
C. Topics for Future Research ……..…… 68

LIST OF REFERENCES …… 73

INITIAL DISTRIBUTION LIST ……75

LIST OF TABLES

1. United States Army Training Management Cycle……24
2. Comparison of Different Models of Training Evaluation…….69
3. A Model of Evaluation……..71

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENT

The author would like to thank the thesis advisors for their support, advice, and patience. The author would also like to thank NYC Fire Commissioner Nicholas Scoppetta and Chief of Department Frank P. Cruthers for their recommendation and support during the eighteen months of this Master's Degree Program. Most of all I would like to thank my wife for her support, sacrifice, and her confidence in me.

THIS PAGE INTENTIONALLY LEFT BLANK

I: INTRODUCTION

A. PURPOSE

The Fire Service is at an important crossroads in its history. The increasing threat of terrorism, along with the need for the Fire Service to make a major contribution to the Homeland Security effort compels firefighters to expand upon their traditional mission requirements. They must deepen and widen the skills they possess for responding to all-hazard[1] incidents, so that they can be ready to apply those skills to the terroristic events[2] and weapons of mass destruction incidents that are inevitable during this war on terrorism. This will only be possible if the Fire Service embraces major conceptual changes in the way that it trains and exercises its members. In many respects, that exercise and training system is excellent and serves as a model for other disciplines. In other ways, however, it tends to be flawed and burdensome, preventing the Fire Service from taking major strides forward in its training efforts – especially with respect to the use of metrics to objectively evaluate performance capabilities for Homeland Security-related operations, and the embedding of assessment techniques in a broader system to provide for improved performance.

This thesis examines the current approach taken by the Fire Service to training and exercising, especially in the Fire Department of New York (FDNY), and identifies gaps and problems in those existing systems. Then, the thesis examines relevant "best practices" from the United States Military and the private sector that might be applied to the shortfalls in Fire Service training and exercising. The remainder of the thesis proposes changes to the training and exercise system, designed to make a measurable, sustained impact on the capabilities of the fire service to accomplish Homeland Security and traditional mission requirements.

[1] An all-hazards event is defined, for the purposes of this thesis, as common fires and emergencies that the first responder community responds to on a daily bases to protect the life and property of the citizens that they protect. They can range from fire and medical emergencies to building collapses, flooding and hurricanes, anything that puts the public at large in danger.

[2] A terrorist event is defined, for the purpose of this thesis, as an event that involves a CBRNE agent or precursor that was intentionally placed to do harm to a person or object with the desire to influence change.

B. THESIS ARGUMENTS

This thesis will argue that while the Fire Service has pioneered the practice of training in harsh, realistic conditions, and has developed an excellent evaluation process to assess the performance and attainment of minimum standards by new recruits ("Probies"[3]), the broader training and exercise program suffers from major flaws. Compared to the U.S. military and the private sector, the fire Service has made very little progress in the use of metrics to assess performance in exercises, and to utilize those performance metrics as part of a sustained system to improve Fire Service capabilities. This is a huge opportunity for improvement. After having completed the research for this thesis, I am a firm believer that "you can't improve what you don't measure" (Harbour, 1997). Indeed, the ability of the Fire Service to achieve its strategic plan for the future, and be the contributor that it must be for Homeland Security, is dependent upon the development and application of appropriate metrics in the exercise system.

The need for progress is especially glaring in the realm of multi-agency exercises. In Homeland Security, multiple agencies (police, fire, public health, etc) must be able to work together in response to a terrorist strike. Exercises should be structured to objectively assess their ability to do so, within a broader system that would then provide for the improvement of capabilities that the exercise indicated were inadequate. Little of the sort exists today. Too often, exercises have fallen victim to the wishes of the participating organizations (and their leaders) to look good. Nobody wants to be perceived as an unprepared or incompetent organization. Hence, in many exercises (multi-jurisdictional, multi-agency or local), serious assessments of performance never happen. All the agencies hold a group hug, sing kum-by-yah, slap each other on the rear end and go home, complacent in their mediocrity. There is no attempt at real institutional change or improvement of preparedness levels. The Fire Service can, and must, do better than this.

The Fire Service is also faced with the potential of becoming a major player or a minor participant in the consequence management of hazardous incidents. There are entities in this widening playing field of terrorism response that are poised to profess expertise and dominance over the operational world of consequence management, once

[3] Probie is the name the FDNY calls a first year firefighter. It is short for probationary firefighter.

the sole possession of the Fire Service. It (the Fire Service) must confront this challenge and re-exert its capability and proficiency that exists because of its strenuous training ideology that it professes and implements. It must prove its ability to continue to be the primary force, not through rhetoric, but through action, the actions of superior proficiency in the essential tasks that are needed to accomplish their mission. They must continue their efforts for operational and equipment competency, that is enhanced by effective training. They must elevate their preparedness assessment that will result from efficient production and measurement of operational exercises. And they must always profess their readiness that will result from the accumulative effects of all the former.

The Fire Service must use this opportunity to revitalize and strengthen the value of exercises by leveraging federal grant funds to build an exercise system that will effectively reflect their capability and superiority over other organizations that covet their missions. They must, through this revitalized exercise program, establish their preparedness and solidify their rightful position as a major contributor to Homeland Security. This can only be accomplished if they develop proper metrics to measure their accomplishments and the accomplishments of rival agencies so that their expertise and competence will show above all pretenders. In turn, competency will make policy makers take notice. The war on terrorism and local preparedness is a joint effort, not the sole domain of a single agency.

The U.S. military and the private sector have developed better practices for exercises and training that are potentially of great value to the Fire Service. While the Fire Service has its own unique requirements and characteristics, and no one agency can copy the exercise system from another, the research of the literature indicates that a number of key improvements can be made if the Fire Service draws from the best practices in industry and the military. These improvements will enable fire chiefs to assess the capability, preparedness and readiness levels of the units within their battalions and their departments as a whole. They will also allow performance comparisons and assessments between disciplines when used in multi-agency exercise scenarios. In particular, this thesis proposes the adoption of a metric system that could be applied to the training and exercise programs of first responders, specifically the Fire Service, to

afford the leadership of these organizations the ability to assess the preparedness of their members to manage the consequence phase of an all-hazards or a terrorist event.

C. RECOMMENDATIONS OF THE THESIS

The recommendations that will be presented in this thesis will result in a significant advantage for the Fire Service to predict and assess its preparedness over other organizations.

The first recommendation is that FDNY and all other participating agencies in the city-wide exercise program accelerate the development of an Exercise Design Teams (EDT) as called for by the Homeland Security Exercise and Evaluation Program (HSEEP). Each agency should also develop an Executive Committee for Preparedness Evaluation (ECPE). This would enhance the coordination and cooperation among agencies and establish a single point of contact and accountability within the agency.

The next recommendation would be to design exercises that are simple and focused to specific goals and objectives. Many exercises today cover too many diffuse objectives. The EDT should restructure exercises so that they concentrate on the key competencies and objectives of FDNY's strategic plans. Scenarios should be realistic, and the performance measures for each exercise should be specified with much greater clarity. This will also help facilitate comparisons among exercises and agencies.

Exercises that are multi-jurisdiction or multi-agency should stress the need to follow an appropriate Incident Management System (IMS), and the agencies should be held accountable for improving any deficiencies in performance.

Accomplishing these objectives will require that the Fire Service answer the "who, what, where, why and how" question. The answer to the why question (i.e., why should we restructure the fire service's approach to exercising) – is clear: building a metrics-driven exercise system is essential to taking the already excellent capabilities of the Fire Service to the next level, so that the Service can be a first-rate contributor to Homeland Security (and be even better than it is today in performing traditional missions). The other questions are more difficult to answer, and provide the primary research focus of this thesis. Who exactly should be exercised (ranging from performance assessments of individuals through units, battalions and divisions)? What skills should be

measured? When and how can those assessments be made in an objective way, as opposed to the current reliance on subjective criteria (or no real criteria at all) prevalent in too many exercises today?

It is further recommended that objective performance criteria be developed for each of the critical tasks that firefighters must accomplish within a given mission. Methodologies for developing these metrics can be derived from a number of models in industry and the military, and from the sizeable academic literature which has emerged on the topic (which is reviewed and critiqued in chapter 3). The selection and application of those metrics should reflect the recognition that more criteria are not necessarily better. The Fire Service should measure only the aspects of an exercise that is pertinent to the goals and objectives of that particular exercise and scenario. These goals and objectives should reflect the broader strategic plan of the organization, rather than a vague sense of what is desirable.

In applying these metrics (the "when" question), the Fire Service should adopt a "before, during and after" approach. That is, the service should 1) assess the capabilities of an individual or unit before training or exercising; 2) assess performance during the exercise; and 3) measure the performance after the exercises (and after any remedial training programs) are concluded so that commanders have an accurate assessment of overall capabilities.

As in the military, comprehensive After Action Reviews (AAR) must become standard components for exercises, so that good practices can be reinforced while deficiencies are identified and corrected.

Most important, the desire of stakeholders to make their organizations look good in exercises cannot be allowed to corrupt the exercise process, or weaken the value of exercises to identify performance gaps and drive performance improvements.

The military commander is responsible to assess the capabilities of his command so that preparedness and readiness can be predicted. To help make improvements in this regard, it is recommended that the Fire Service borrow from this military "best practice" in how assessments are reported and used within the organizations. The military often involves very senior levels of the organization in the training process, by training one level down and evaluating two levels up. The Fire Service hierarchy is generally far

more distant from equivalent training activities, weakening their ability to use the system as an effective tool to understand and improve service capabilities. Adopting a system for closer leadership involvement would be of significant benefit.

D. SCOPE OF RESEARCH

This thesis includes the review of the current exercise system that is being used throughout the first responder community and is supported by the Department of Homeland Security, Office of Domestic Preparedness[4]. This review is done in order to show the types of exercises in use today to establish a baseline to discuss improvements that can be accomplished.

This thesis will review the United States Army Training Manual, Training the Force[5], and focus on their assessment strategy to evaluate performance and assess combat readiness. This will ground our recommendations in a proven practice of evaluation, feedback, and assessment. Though, not perfect, the military system has valuable concepts to offer this discussion.

This thesis will acquaint the reader with various performance evaluation models that are in use in the industrial sector to evaluate a corporation and its performance on many different levels from personnel and production processes to financial stability and provide a comparison chart for eight of the most popular models. This will enable the Fire Service to adopt proven evaluation practices that will result in heightened preparedness and readiness.

Also this thesis will analyze the training program of the FDNY and discuss some of the challenges that it faces now and in the future. This thesis will also outline criteria for measuring performance that should guide the creation of metrics to validate preparedness. Finally, this thesis will propose a metric system that will allow the

[4] The Department of Homeland Security, Office for Domestic Preparedness, has published, Homeland Security Exercise and Evaluation Program, which serves as the National Exercise Program guidance document.

[5] This document, the United States Army's, Training Manual, Training the Force, serves as the overarching training manual for all the military services and treats training and operations as a combined effort of all entities to complete the task. This serves as a great example for a multi-agency, multi-jurisdictional combined effort at terroristic activity. No one agency is the end all and be all for crisis and consequence operations.

leadership of the organization make an accurate assessment of the capability of the first responder standing force.

E. RESEARCH METHODOLOGY AND THESIS STRUCTURE

The research methodology begins with a comprehensive review of literature relating to performance measurement, strategic planning, evaluations and metric methodologies. The review of literature included books, professional journals, Department of Defense publications, Department of Homeland Security publications, Congressional legislation, Fire Department of New York City policies and procedures, and private internet sources. The best practices of the organizations involved in the literature will be extracted and used to recommend a metric system to evaluate and assess the preparedness of first responder units and organizations.

The thesis begins by reviewing the current exercise system of the Fire Service in Chapter II (and, more specifically, FDNY), and assessing how that system is evolving in response to the new requirements driven by Homeland Security. Chapter II also analyzes the implications of emerging Federal requirements on the exercise system for FDNY efforts to improve its exercise system. The analysis identifies a number of excellent practices in the Fire Service that need to be extended throughout its exercise system, and also significant new opportunities to build metrics into FDNY's exercise evaluation system. Chapter III examines best practices from the U.S. military and from the private sector that might be applied to the Fire Service, and reviews the academic literature on methodologies to build metrics and performance measurement systems. Chapter IV applies lessons learned from the review of the literature and best practices to the development of performance criteria for use by the Fire Service. Chapter V proposes how those metrics might be incorporated into a broader restructuring of the Fire Service's exercise system, with Chapter VI offering my conclusions.

THIS PAGE INTENTIONALLY LEFT BLANK

II: THE CURRENT EXERCISE SYSTEM OF THE FIRE SERVICE: EXCELLENCE AND OPPORTUNITIES FOR IMPROVEMENT

A. INTRODUCTION

> As first responders to fires, public safety and medical emergencies, disasters and terrorist acts, the FDNY protects the lives and property of New York residents and visitors. The Department advances public safety through its fire prevention, investigation and education programs. The timely delivery of these services enables the FDNY to make significant contributions to the safety of New York City and homeland security efforts.
> Core values: Service, Bravery, Safety, Honor Dedication, and Preparedness.
>
> -The Mission Statement of the Fire Department of New York as stated in their Strategic Plan of 2004-20, dated January 1, 2004

The realities of the post 9/11 world have sown the seeds of change in the Fire Service. The Fire Department of New York City (FDNY) has begun this process of change by redefining its mission statement, restating its core values and publishing its first-ever strategic plan to guide the Department in its effort to meet its new responsibilities in the war on terrorism. The Department is not only following this new strategic plan, but more importantly, the leadership and the members are now thinking strategically about how to build its capabilities to meet the requirements of Homeland Security.

Some features of the current exercise system are well suited to strengthening Fire Service capabilities (as will be demonstrated later in this chapter). Overall, however, the Fire Service can not validate its preparedness for terrorism response because it uses a system that is based in observational evaluation of capabilities, which tends to be subjective and potentially biased. The fire service has long believed that training is a continuous, career long endeavor that, if accomplished correctly, will produce competent, confident, disciplined and adaptive firefighters and officers. It was an easy methodology for the fire service to follow in the past. We could train firefighters cheaply and efficiently through on the job experience with a heavy workload and plenty of experienced veterans willing to share their accumulated knowledge. In the post 9/11 world, this is no longer the case. The Fire Service has entered into a new era of requirements for performance standards, competencies, evaluations and assessments.

The Fire Service has embraced and adopted the "National Exercise Program"[6] (NEP) as part of their efforts to evaluate the performance and competency of its members. They are now faced with the opportunity to restructure their whole approach to training that will enable them to become an effective contributor in the war against terrorism. This can happen by using the exercise as a final evaluation and assessment tool to validate the training program. The adoption of metrics within the exercise system and the whole training program will enable the Fire Service to measure their performance, identify gaps within the training system, test new procedures and plans for effectiveness, and provide an assessment of the organization's capability. Metrics will enable the Fire Service to establish a foundation upon which knowledge gained through experience can be transferred to new learners through the learning cycle, and thereby reinforce a sustained effort to build the new capabilities at the unit level.

The fire service is not alone in facing this challenge. The technological age (the late 1900's into the 21st century) has brought with it an explosion of knowledge that feeds a technological boom world-wide. We no sooner master a new technology than it is replaced by a more advanced one that requires additional skills and knowledge. Performance and performance measurement becomes the underlying indicator that an institution has successfully achieved knowledge continuity and growth. Many researchers have studied these phenomena and examined the ways that we learn and transfer knowledge. The one thing they all agree upon is that in order to know where we must go (that is, how performance needs to be increased), we must remember where we came from (our past performance level) and where we are at the present time so that a systematic way of improving performance can be structured. Each step along the way should require a metric or set of metrics to guide us along the path. Moreover, every organization should have a strategic plan with goals and objectives that are clearly discernable, with performance metrics embedded in an overall exercise and training system that can support the implementation and achievement of that plan. The training that an organization develops must lead to the achievement of the organizations objectives and goals. The only way that an organization will know if it has met its goals

[6] The National Exercise Program (NEP) has become the national standard to which exercises are compared. The program was created by the Office for Domestic Preparedness and has established the HSEEP, Homeland Security Exercise and Evaluation Program document as the cornerstone of the program.

and adhered to its strategic plan is through the metrics that it creates to measure its past, present and future capabilities.

This Chapter will examine how effectively the fire service is meeting these overall requirements for effective performance measurements and drill down into some specific opportunities for improvement. The chapter begins by reviewing some key sources of Federal guidance in restructuring exercises for homeland security, and summarizes the key features of the existing training and exercise system. This chapter then examines emerging challenges that the system confronts, especially those related to terrorism response. After reviewing some important new initiatives designed to meet these challenges, the chapter concludes by highlighting some remaining requirements for change.

B. THE HOMELAND SECURITY EXERCISE AND EVALUATION PROGRAM (HSEEP)

1. Overview of the HSEEP

> Past experience has shown that preparedness efforts are key to providing an effective response to major terrorist incidents and natural disasters. Therefore, we need a comprehensive national system to bring together and command all necessary response assets quickly and effectively. We must equip, train and exercise many different response units to mobilize for any emergency without warning.
>
> - National Strategy for Homeland Security, July 2002

The Department of Homeland Security, Office for Domestic Preparedness (ODP)[7], has implemented a program of financial and direct support designed to assist state and local governments with the development and implementation of an exercise and evaluation program to assess and enhance domestic preparedness. This program provides the blueprint for the Fire Service to produce and perform an exercise and provides assessment tools to gauge preparedness. The program proposes a cycle of exercises of increasing complexity, two tabletops and an operational exercise that are designed to test performance as well as the adequacy of plans, policies and procedures. The scenarios that the program suggests are threat-based, realistic, and generally involve multi-agencies and

[7] U.S. Department of Homeland Security, Office for Domestic Preparedness, Homeland Security Exercise and Evaluation Program (HSEEP),is available at www.ojp.usdoj.gov/odp, September 2003.

jurisdictions. All the exercises are evaluated and the host organizations are provided with After Action Reports and Corrective Action Plans. Lessons learned and best practices from the exercises are recorded and shared with the homeland security community.

The evaluation plan for the program is broken down into two parts, evaluation of tabletop exercises and evaluation of operational field-type exercises. The Tabletop evaluations consist of the activity and discussions that goes on during the exercise which is recorded, analyzed and formatted in a summary report.

The evaluation of the operational field-type exercise consists of an analysis of individual performance, team/discipline/department-level performance, and a community/mission-level performance. The performances are observed and recorded, and the players are given an opportunity during a hot-wash to discuss the actions and performance of their respective mission areas. The After Action Report (AAR) is the crucial document that documents performance and captures the analysis of the observers. It provides a description of events, issues that arose and how they were handled, and recommendations for corrective actions. The final piece to the evaluation system is the Corrective Action Plan (CAP) that recommends the changes that need to be made to the plans, policies and procedures of the organizations.

The HSEEP outlines an evaluation process that uses objective observers to record the performance of the players and the issues that may arise during play. Many times the observers are from the very organizations participating in the exercise which may lead to biases arising in the final analysis of the AAR and the CAP. This is a major problem that needs to be remedied. Organizations cannot be relied upon to evaluate themselves in the absence of genuine, objective metrics and a system to minimize the inevitable desire of organizations “to make themselves look good.” If an exercise is going to be used as an assessment tool for a department or a municipality, then it must be independent, no-notice, fully evaluated and measured through objective criteria. The results must also be verified multiple times over a period of time that is appropriate for the organization. Moreover, while the AAR and the CAP can be transformed into good tools for the Fire Service to adopt into their programs, the exercises must be tied back to training organizations to complete the training cycle. The HSEEP falls short in specifying how to transfer knowledge and learning back to training, so that performance shortfalls

can be remedied in a structured, consistent way. In sum: the HSEEP is valuable, but the Fire Service needs to move beyond it in building a metrics-driven exercise system.

C. CURRENT FDNY TRAINING AND EXERCISE MODELS

The FDNY uses exercises as a means to accomplish hands-on training, to reinforce training objectives and goals, and as a means to validate training courses. Exercises are a standard at the end of firematic and hazmat courses to put learned skills to the test and to integrate the many skills required to complete a mission. Most of these exercises are realistic scenario-based exercises, and are observed and evaluated by the instructor to correct any deficiencies and to summarize the learning. A well designed and executed exercise is also an effective means to:

- Test policies, plans, and procedures
- Clarify and train personnel in roles and responsibilities
- Identify gaps in resources
- Improve individual performance
- Identify opportunities for improvement
- Improve interagency coordination and communications

The FDNY is striving to improve its exercise model to mirror the elements present in the HSEEP model and has secured funding to establish an Exercise Design Team. This team will be able to provide assistance to Battalion and Division commanders who wish to produce an exercise for their command so that they can evaluate capabilities, preparedness and readiness to accomplish specific missions within their jurisdictions.

This Exercise Design Team (EDT) emulates the proposed team contained within the HSEEP but, as opposed to a transitory team, it would be a permanent unit that will act as a liaison for the small, fire department exercises as well as multi-agency, multi-jurisdictional exercises. The EDT will standardize the exercise process and pull the best practices and evaluations from the NEP, military and industrial models.

In terms of training, some of the very best practices in the Fire Service are reflected in the Probationary Firefighter School Model – practices that might be applied much more widely by the EDT to other exercise and training programs in the service. The motto that guides the training attitude of the FDNY is emblazoned above the

entrance to the Probie equipment room: "Let no man's ghost return to say his training let him down".[8] The standards that are set by this one motto are high, and instructors push the limit to enable Probies to meet that standard. The knowledge that is learned, the skills that are honed and the competencies that are achieved start the new firefighters on their way in an effective fashion. It is the combination of what is learned and measured in the initial stages at the academy that will stay with them their whole career.[9] The Probationary Firefighter School (PFS) uses a semi-military disciplinary style that creates a structured regiment of learning. The probie, as they are called, is introduced to fire tactics, fire prevention, emergency medical procedures and hazardous materials operations over a twelve week period.

Academically, performance is measured and tracked using a series of weekly quizzes, a midterm and final exam, and they are also tightly monitored over their twelve week period at the school. The academic performance is closely watched and the probie is given every opportunity to improve through individual training, peer training or extra sessions if the need arises.

The hands-on training (or "train as you fight"),[10] accounts for 60% of the training skills that are taught and measured. This is evaluated by their instructor, either by observation that a skill is performed or with reference to a checklist of steps in a properly performed evolution, forty-one times over a twelve week period. If any evaluation is unsatisfactory or the probie fails to perform adequately, the problem is addressed forth-with using an intimidation tactic that places the probie in the office of the Captain-in-Charge of the school for a 'very personal' interview.

This model works well for many reasons. It highlights the seriousness of the career that the probie is about to enter. It strives to prepare them for the life or death situations that they will find themselves. Every tactic and procedure in the twelve weeks of training prepares them for the reality of firefighting. They train as they fight, with the same equipment, in realistic environments, and with the same high stress. Every instructor strives to place the probie inside the organized chaos that is the fire floor. The

[8] Unknown author
[9] This information comes from an interview with Lt. John Regan, the former Executive Officer of the Probationary Firefighting School, NYC Fire Academy, Randall's Island, NY, 10035, and the author, April 29, 2004.
[10] Train as you fight is the motto of the military training manual,

training and evaluation process instills a sense of discipline and respect for chain of command that exists in the fire department. And finally, the PFS model instills within the probies the ability and the necessity to take an order without question.

The final evaluation of all the new probies comes from the Chief-in-Charge of Training who has been actively evaluating and observing the formation of these skill levels. All the evaluation reports are at his disposal to formulate his assessment of their preparedness to join the active force. If a probie fails, they will have to repeat the twelve weeks or be terminated from the force. This is a tough decision but one that has to be enforced for the good of the service

Other models (and potential best practices on which to draw more widely) exist in specialized training programs. The CFRD[11] (Certified First Responder, Defibrillator) Certification Model is valuable in this regard. This training in medical emergency procedures is strictly monitored by the Fire Department Medical Affairs Division and is evaluated on a regular basis. The students are tested academically through quizzes and examinations and must pass a New York State mandated exam before being certified. There is also a hands-on skills workshop that is observed and recorded on a check-off list basis. The evaluation is quantified to produce a performance profile on ability. The course instills an environment of seriousness, capability evaluation, accountability, and leadership involvement. This effort produces qualified CFRD firefighters that are capable and confident in their approach to their work.

Hazardous Materials Training, both Operations and Technician[12] Levels, provide another model of how objective criteria can be brought to bear in performance evaluations. This was developed and is sponsored by the International Association of Firefighters (IAFF) through a grant process from federal sources. The program strives to deliver a course that meets the highest standards for efficiency and effectiveness for the audience they are addressing (namely, the adult first responder). A rigorous Quality Assurance Program is employed to generate a report containing selected performance

[11] CFRD or Certified First Responder, Defibrillator, is a New York State category especially created for first responders in NYC to fill the need to have certified personnel placed quickly on the scene of a cardiac emergency or other trauma. They yield patient care to Paramedics or EMT's upon their arrival. CFRD is a level below an EMT (Emergency Medical Technician) who works on a fire apparatus instead of an ambulance.

[12] The Code of Federal Regulations, OSHA, CFR 1910.120, describes four levels of training for Hazardous Materials workers, Awareness, Operations, Technician, and Incident Commander.

indicators that are coupled with pre- and post- test scores that are used to represent a percent knowledge gained. A third test is given to selected students one year after the course to measure the amount of knowledge retained. This program of Quality Assurance, though not new to other disciplines, is an innovative approach in the Fire Service that has proven success over the years.

Numerous other approaches to training are used at the Fire Academy because of the cross section of experience and seniority among the members trained. The most popular model mirrors some of the best practices of the military evaluation program that uses instantaneous feedback and observation as their assessment tools. The Probationary Firefighter School uses this model and expands upon it with great success. Most other courses used for in service training of new procedures (or little used competencies which train more senior personnel) use observation of hands-on evolutions as their evaluation methods, with a training mark given at the end of the session. End of course reaction questionnaires or "reactionaires"[13] (as they have come to be called) are used in a limited fashion to evaluate instructors, course content and course environment.

D. CHALLENGES FACED BY THE FDNY AND THE FIRE SERVICE

There are many challenges that exist for the Fire Service and the FDNY in their training routines. The main challenge exists in the department's ability to take capabilities built from the training process and translate them into an accurate preparedness model. The Fire Service needs to adhere to the standards[14] of OSHA 1910.120 and NFPA 472 as a minimum base line for members operating in hazardous environments and at terror events, build upon that foundation with new standards that will raise the bar of preparedness.

The largest, single challenge that faces the FDNY and the Fire Service is the maintenance of competency in skill sets that are seldom used but are critical to complete the missions for which they were designed. Members remain competent in many of the skills used in firefighting and medical care because they use those skills daily in their

[13] Reactionaires is a coined phrase for questionnaire by the IES in their study by Tamkin, et, al., p 31.
[14] OSHA 1910.120 and NFPA 472 are the standards currently recognized as the guiding influence when training hazmat personnel to deal with hazardous material incidents. They deal with operational principles and personal protective equipment usage.

responses which are reinforced in their experience gain. The multitudes of other skill sets that are used in conjunction with these often used skills are practiced in the firehouse so that they are at a high readiness posture.

The war on terrorism has mandated that the Fire Service learn an additional set of skills that are not readily experienced in the field. We can only train in a simulated environment. This presents a large challenge to the training professional in the Fire Service to keep the standards high while maintaining the interest and determination of the adult, experienced, senior firefighter.

The influx of numerous detection devices specific for terroristic activity, and the additional skill sets needed to evaluate these devices, has created another large problem in a force that gains knowledge through experience (even if simulated). The experiences are simply not there. New learning and experience techniques need to be developed to artificially produce skill sets that are adaptive to the new required skills. These skills must also be measured and recorded to improve our posture.

The training model of the FDNY is typical of the Fire Service in general and needs to increase its flexibility and its dynamics. The mission that the Fire Service has undertaken is different from the past, and requires new ideas and modern technology to meet the rising challenges. The leadership must be proactive in its thinking and actions. Evaluation of training in order to evolve is tantamount to an effective, capable, prepared and ready force.

Single-agency exercises are the capping stone that should accompany any meaningful training regiment. They validate the training process and if measured correctly, will give the commander a picture of the capability, preparedness and readiness of the command. The biggest challenge of the Fire Service is to provide a realistic, meaningful exercise that will challenge the training of all of its members and supply the proper measures to analyze the outcomes.

Presently, exercises run within the department, for the department, are specific training exercises that use observation by the commanders to evaluate the outputs. They are generally successful efforts participated in by motivated personnel that genuinely produce effective results in knowledge transfer from members to members, commanders to members and even members to commanders. At the conclusion of the exercise,

whether it is a small end of course evaluation, hands-on exercise or a large operational Hi-Rise exercise, the favorite evaluational tool is observation-based critique along with an After Action Discussion forum.

Inter-agency exercises are becoming a larger challenge for the department, especially with regard to terrorism response. Here, operational lines are blurred and mission areas are not defined. Measurable metrics are frowned upon because they may point to a deficiency in cooperation (which would be politically embarrassing). The Incident Commander, who may have a relatively accurate idea of the capabilities and preparedness of his members, has no effective assessment of the capabilities of other agencies. In the past, this has led to dangerous situations for the operational forces at actual incidents. Many gaps in procedures from one organization to the next become apparent, but neither organization is willing to admit to its own shortcomings. Unlike the system called for in National Incident Management System, there is no single manual that describes procedures for joint operations that is accepted by all players in NYC.

Herein lies the challenge for the FDNY and the City. The fact is that at any future terrorist event that includes a consequence management response (traditionally a Fire/EMS responsibility), that response will also include the response of other agencies that will cause operational competition, a competition on a Command level as well as at an operational level that will lead to command and control problems.

E. EMERGING BEST PRACTICES

The FDNY is headed in the right direction in establishing the Exercise Design Team as an institutional means of coordinating and standardizing exercises without dictating content in an overly detailed way. In many other respects as well, initiatives are underway in FDNY that – as part of a broader restructuring of the exercise system—should be incorporated in that redesign process. Some of the best practices are as follows:

- The local division and Battalion Chiefs have the ability to produce training exercises specific to their area.
- Chief Officers often take significant roles in the production of these single agency exercises.
- Policies and procedures are put to the test and evaluated.

- At the Fire Academy, members are trained in regularly-scheduled, hands-on exercises that strive to reinforce training across the service.
- Exercises typically use realistic scenarios and equipment.

A new best practice is emerging with regard to the special challenges of the multi-agency operations and exercises. This has been termed the Phased Response Strategy Exercise[15] and is being developed by the FDNY 1st Division, in lower Manhattan. It is based on the assumption that every incident has a logical flow that can be put into phases: the incipient or response phase, the mitigation phase and the overhaul and investigative phase. In the field, this is a pattern that real incidents follow and because of reality and the actual response patterns of the responding disciplines inter-agency problems seem to be non-existent or just melt away. Exercises, in contrast, are different due to the sterilized environment in which the scenario occurs. Every response organization is unrealistically placed at the scene, if not immediately, then close to it, where all their resources and assets can be placed in play, which leads to disruptions and competition.

In the Phased Response Strategy, agencies work on independent problems in phases as they would during a real incident, each agency completing their assigned missions. For example, a bomb is found in Penn Station, the major organizations that would respond in phase one to assess the incident are: FDNY, NYPD, FBI, and DEP. They would logically and in reality not respond at the same time. Initial units on the scene would be FDNY and NYPD. Since it is a bomb (phase two begins) the law enforcement and bomb squad would initially operate and FD would be in a support role. Events are similar in the exercise. The two organizations are completing their primary missions. In the scenario the bomb detonates causing collapse and fires (phase three). The PD exits to regroup and the FD enter to complete fire suppression, assessment, rescue and collapse shoring. EMS begins play for victim assessment, triage and transport. As phase three missions are completed, phase four begins with investigation and overhaul. As in reality, there is a logical movement through specific missions for each discipline. Any number of activities can be added for the specific organization. Overlap of phases can be

[15] The Phased Response Strategy is the brain child of a Battalion Chief Bill Hines of the FDNY Battalion 7, Division 1, with his approach of making exercises beneficial in Penn Station when operating in a multi-agency environment.

implemented within the exercise but the evaluation, performance measurements and objectives of each agency must be obtained separately. Exercises today tend to press all activities into one short time frame. They create evaluation problems along with cultural tensions.

The Phased Response Strategy elongates the exercise process so that there can be an exchange of knowledge within units and between agencies. Each move can be analyzed as it happens and recorded. It is experience that is learned by doing and repeating. The present exercise system speeds through the operations phase so that they can achieve the outcomes before all the inputs and outputs are accomplished. That is the fallacy of the present system.

F. SHORTFALLS

Despite these important elements of progress, the FDNY has some improvements that still need to be made to its exercise program. One problem is that many exercises try to accomplish too many objectives simultaneously. Learning and institutional capacity-building is a grassroots phenomena that starts at the individual level and works its way horizontally and vertically. Individuals learn and absorb knowledge. Individuals of an organization learn and absorb knowledge. Organizations can only learn and absorb knowledge because of the individuals that are within the organization. The main reason that major exercises fail is that they try to measure and evaluate everything that is happening within that exercise in a short period of time. The metrics employed fall short of diagnosing an institutional illness or identifying a correctable problem. The players on a local level become disenchanted when large numbers of other agencies show up and monopolize the activities. This throws off the metrics of the individual and team.

Exercises and exercise evaluation needs to have a tiered approach throughout training. Smaller exercises evaluate the small unit responses and their competency. Larger exercises evaluate the command and control aspects of the operation. If the units have proven their competency in smaller exercises, then the factors that influence the larger exercises and push performance will be the command and control and that is the competencies that should be measured. Likewise, if an agency has proven its ability to properly used incident management skills and successfully operate at an incident, then at

a multi-jurisdictional incident, the emphasis should be on the ability of agencies to operate as a single force and the effects this command has upon the operational forces.

An additional problem lies in the end of course reaction questionnaires or "reactionaires" (as they have come to be called) which are used in many exercises. They are intended to be used as a means to evaluate instructors, course content and course environment, but are often of little value for measuring performance. They do not include the kinds of specific, objective metrics that can be used to effectively measure performance. At present, however, the necessary system does not exist to develop and evaluate such exercise metrics. The fire service also needs to establish a standardized reporting system for exercises that quickly provides reports to both the Chief Officers and members, so that they can evaluate unit performance and monitor the process by which shortfalls in performance are remedied. At present, there is a general lack of accountability for both inter-departmental and intra-departmental exercises. Who is accountable to train or retrain members? Accountability needs to be more clearly defined.

Exercises are an apparatus to identify the gaps in training, but no reliable mechanism exists to change training when performance gaps are identified. The Bureau of Training needs to be involved outside of the training academy in building such a mechanism. There is a definite disconnect between training and the operational forces in the field. The Bureau of Training does a great job in making much of their training as realistic as possible, even using real smoke and fire for some of their evolutions. This does not get transported over to exercises which still use artificial conditions, and which allow other disciplines and organizations into areas that in reality they would not be able to venture because of the lack of personal protective gear.

Finally, exercises are always planned, orchestrated events in which participants rehearse their roles prior to the event. With the rehearsals, the unrealistic props, the shortened response times and the competition for dominance between agencies, exercises currently provide a poor basis to evaluate capabilities and performance levels.

THIS PAGE INTENTIONALLY LEFT BLANK

III: REVIEW OF BEST PRACTICES IN THE U.S. MILITARY, THE PRIVATE SECTOR, AND AS DESCRIBED IN ACADEMIC LITERATURE

A. INTRODUCTION

The Fire Service is a unique organization that has unique requirements and characteristics. In order to fill the shortfalls identified in Chapter 2, there are important lessons that can be learned and adopted from other organizations that have similar missions and have addressed similar shortfalls. The opportunity exists to take advantage of the lessons learned from the military and industry in their efforts to improve performance and couple those practices with the best practices developed within the fire service itself.

B. THE UNITED STATES ARMY'S EVALUATION AND ASSESSMENT MODEL

1. Overview of the Model

The United States Army is much like the Fire Service. They continuously train to be prepared for the battle. When called, the army battles an enemy in a hazardous, wartime environment and the Fire Service battles the 'red devil', fire, and a host of other hazardous materials also in a hazardous environment. When it comes to terrorism and the threat of WMD, both organizations are trained to be proficient in procedures and instrumentation that, hopefully, they will never have to use. This is the hardest part of preparedness; maintaining competency and proficiency at tasks that are not everyday skills. The only way they will know what level they are at or where they need to go is if they measure their performance and analyze the results. Therefore, the U.S. Army and their training model will serve as one of the foundations for creating metrics that will be useful for the fire service.

The Army trains as they fight, as realistic as they can possibly attain. They assess every aspect of their organization so that their commander can judge their ability to accomplish their wartime operational mission.[16] The Army evaluates, assesses, reviews,

[16] This becomes the first best practice element for the Fire Service, the commander is responsible to assess the capability of his command through evaluations from trainers, company commanders, and his own knowledge and experience.

so that they can plan and execute their mission. There is feedback gathered throughout the process that makes it to the top supervisors so that institutional change can be accomplished if needed.[17]

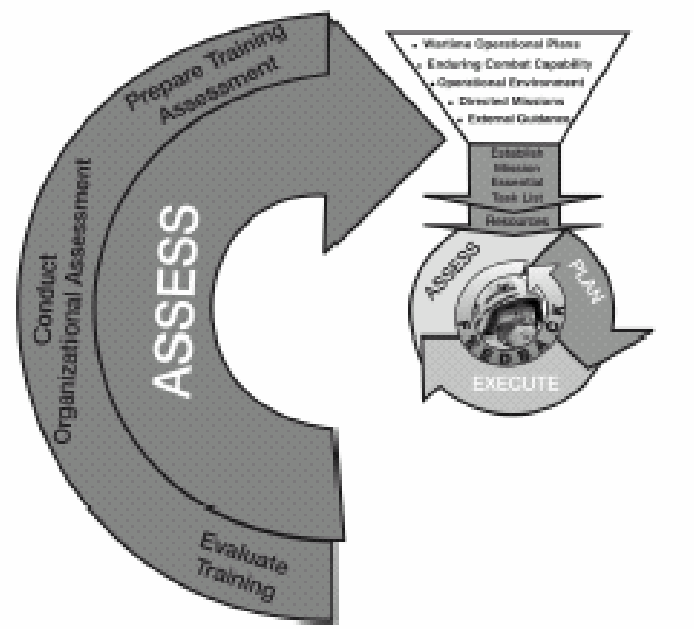


Figure 2.1: U.S. Army Training Management Cycle[18]

2. Assessment and Evaluation

The Army treats everything they do as a training experience and strives to evaluate and assess with the final objective to be able to accomplish their mission and the goal of preparedness 24 hours a day, 7 days a week, 365 days a year. The commander has the responsibility for the assessment of his troops by using his/her experience, knowledge of his troops, personal observation, the feedback from the training evaluation reports from the units under his command, and other reports and evaluations that are feed back for his assessment. This assessment is a continuous process that includes evaluating training, conducting an organizational assessment, and preparing a training assessment

[17] The next best practice is the feedback that is capture in all directions, horizontally and vertically in both directions. This enables the program to adapt to the environment.
[18] Training the Force, p 6-1

which has become institutionalized within the force[19]. The measures of performance are evaluated against the established Army standard to identify gaps in performance and/or training.

It is important at this point to distinguish between evaluations and assessments. Evaluations measure past events in training so that a commander can make a judgment on his commands capability in the future. Assessments are a summary of all the metrics used in the evaluations, along with other measures to predict the future. In this respect, the Army is using performance measures of the past and present to predict the future.

Everyone and everything in the force is evaluated and assessed for operational readiness, from the lowly private, to the powerful general and from the operations of the squad to the logistics of their command and control. Evaluations used by the Army can be[20]:

- Informal – a squad leader training a squad in tactics with immediate, informal feedback. The metrics here are simple and usually verbal.
- Formal – Training and exercises are evaluated by dedicated evaluators and are usually done by two ranks above.
- Internal – are planned, resourced and conducted internal to the group
- External – are planned, resourced and conducted by an echelon outside and above the chain of command

By using assessments and evaluations the way the Army does in its training exercises, they mirror common civilian models that will be discussed in later sections. By evaluating an individual, group, team or unit, the commander evaluates this specific training audience's performance as it relates to the completion of their specific task. When the commander uses the assessment process, he embraces the compilation of training evaluation reports from individual units and groups, feedback from the multiple levels of evaluators and experience from personal observation and knowledge of the training, so that he can predict if his command is able to work jointly to achieve it's objectives. In other words, the commander of the troops ultimately assesses whether or not his command is prepared and what level of preparedness the command has obtained.

[19] Training the Force, Chapter 6, p 6-1
[20] Training the Force, (FM 7-0, Field Manual, 2002),Chapter 6, p6-3

3. After Action Reviews (AAR)

The Army conducts an After Action Review (AAR)[21] after every training event. Feedback is their main assessment tool. The AAR is a method of providing feedback to units by involving participants in the training diagnostic process in order to increase and reinforce learning. The AAR leader guides participants in identifying deficiencies and seeking solutions[22]. The AAR can be formal or informal, structured or unstructured, and, most important of all, it is not a critique. The AAR is the Army's most used and important metric for learning and it has the following advantages over a critique[23]:

- It focuses directly on key METL[24] derived training objectives
- Emphasizes meeting Army standards rather than pronouncing judgment of success or failure
- Uses "leading questions" to encourage participants to self-discover important lessons from the training event
- Allows a large number of individuals and leaders to participate so more of the training can be recalled and more lessons learned can be shared.

The AAR consists of four parts:

1. Review what was supposed to happen (training Plan).
2. Establish what happened
3. Determine what was right or wrong with what happened.
4. Determine how the task should be done differently the next time.

> The key to fighting and winning is an understanding of 'how we train to fight' at every echelon. Training programs must result in demonstrated tactical and technical competence, confidence, and initiative in our soldiers and their leaders. Training will remain the Army's top priority because it is the cornerstone of combat readiness!
>
> -General Carl E. Vuono[25]

[21] Training the Force, Chapter 6, p 6-4
[22] Definition from Training the Force, Glossary, p 8
[23] Training the Force, (FM 7-0, Field Manual, 2002),Chapter 6, p 6-4
[24] METL (mission essential task list) is a compilation of collective mission essential tasks an organization must perform successfully to accomplish its wartime mission(s) [Training the Force, Glossary, p13]. METL's are very similar to competencies used in the fire service.

4. The Army's Ten Principles of Training

The Army model, Training the Force, (FM 7-0, Field Manual, 2002), serves their organization and their objectives well. The Army has Ten Principles of Training[26]:

- Commanders are responsible for training
- NCO's train individuals, crews and small teams
- Train as a combined arms and joint team
- Train for combat proficiency
- Realistic conditions
- Performance-oriented
- Train to Standard Using Appropriate Doctrine
- Train to Adapt
- Train to Maintain and Sustain
- Train Using Multi-echelon Techniques
- Train to Sustain Proficiency
- Train and Develop leaders

5. The U.S. Army's Train the Force Model - Good Points, Bad Points and Its Implications for the Fire Service

The Army's model serves as a template for the entire military. Though they lack actual quantifiable metrics, they do a good job of limiting outside influences on their qualitative metrics. This allows the final assessment of readiness (preparedness for the Fire Service) to be a very objective decision by the commander and it can leave room in the final formula to make very general assumptions. If the commander is an authoritative personality that prefers to make decisions on his own without consultation with qualified peers, the results could be deadly.

The military culture of a supreme hierarchy and obedience to authority limits potentially corruptive cultural influences that the civilian world must endure. Certain biases and poor work ethics have little place in the military thinking and thus do not exert the pressures as they do in the civilian world.

---

[25] This quote can readily be adapted to the new mission that must be performed by the Fire Service.
[26] Training the Force, (FM 7-0, Field Manual, 2002), pp. 2-2 through 2-9.

The concepts that should be mirrored in the fire service when developing metrics for performance are;

- **Evaluation and Assessment -** the use of evaluations to measure the past and present in order to assess the future capabilities. The Fire Service uses evaluations to improve training and improve the performance of the members. A disconnect occurs when these evaluations are not forwarded to the leadership for assessment of the training and eventual assessment of preparedness.
- **Hierarchical Involvement -** the involvement of the highest levels of the organization in the training process. The Fire Service hierarchy generally is to far distant from the actual process. The military practice to train one level down and evaluate two levels up will bring the hierarchy a step closer to the operational building block – the unit.
- **Accountability -** their ability to hold personnel and their leaders at every level accountable for training. The commander[27] is responsible for the training and performance of members but the NCO does the actual training and the NCO is held strictly accountable by the commander. The Fire Service, to the contrast, also makes the company commander of the unit the responsible party for training but is consistently very lax in holding that same commander accountable.
- **Assessment of Readiness by the Commander (only) –** The commander is solely responsible for the assessment of the troops under him in the military model. This practice works well in the military because they can control outside influences on the decision process but is a dangerous practice for the Fire Service. This thesis would support an Executive Committee for Preparedness Evaluation (ECPE) who would have the task of assessing the preparedness levels of units.

[27] Training the Force, (FM 7-0, Field Manual, 2002), p 2-2

C. BEST PRACTICES FROM THE PRIVATE SECTOR

1. Justification for Examining Industrial Models

The Fire Service uses metrics to measure the performance of their members to establish their level of preparedness to react to a situation when called upon to respond. The military uses metrics to measure the performance and capability of their members so that a commander can predict the readiness of the command to complete a mission. So too, the private sector uses metrics to measure every aspect of their performance, from the resource procurement and manufacturing process to the sales and distribution of a product to increase the profit (the bottom line). The motive does not matter, the performance measurement principles are the same for all three organizations. The literature deals more with industry then any other discipline. The common concept throughout the literature is, "you can't improve what you…don't measure".[28]

2. Performance Measurements Models

This section will look at select models for performance measurement that have been developed through the years and vary in small respects according to the emphasis of the particular authors. Table 1[29], taken from a report written by Tamkin, Yarnell and Kerrin, for the Institute for Employment Studies (IES) of the United Kingdom, shows a comparison of eight models that are in use across the private sector to evaluate training and performance within specific organizations. The models are similar but vary according to the specific application they are used to evaluate. Some of the models diverge from the original in their approach to how training evaluation might take place.

a. Kirkpatrick – Four Step Model

> Training evaluation is a bit like eating five portions of fruit and vegetables a day; everyone knows that they are supposed to do it, everyone says they are planning to do better in the future and few people admit to having got it right.
>
> -Tamkin, Yarnell, Kerrin

[28] Harbour, Jerry L., PhD, *The Basics of Performance measurement*, Productivity Press, 1997, p. 1
[29] Tamkin, P., Yarnell, J., Kerrin, M., *Kirkpatrick and Beyond: A review of models of training performance*, IES, 2002, p 13

The first and original model is by Kirkpatrick[30] who is considered to be the father of modern evaluation measurement models because his model forms the basis for all that have followed. In 1959 he described four steps, or levels, as they later became known, that you can evaluate in order to measure whether a training program or an exercise is affective and adhering to the strategic plan of the organization[31]:

- **Level 1: Reaction –** what the participants thought of the training, normally measured in a questionnaire format.
- **Level 2: Learning –** the changes in knowledge, skills, or attitude with respect to the training objectives, normally assessed by use of performance tests.
- **Level 3: Behavior –** changes in job behavior resulting from the program, to identify whether the learning is being applied. Assessment methods include observation and productivity data.
- **Level 4: Results -** the bottom-line contribution of the training program. Methods include measuring costs, quality and return on investment (ROI).

Kirkpatrick's model is simple, easy and sensible as compared to some of the newer models and it is readily adaptable to the fire service to produce performance measures for exercise evaluation. He has been criticized for a too simplistic view that fails to take into account the multiple variables that influence the transfer of knowledge from the starting point on level 1 to the results that actually produce institutional change within the organization because of the training. The metrics that can be applied to this model are simple and easy to use. The questionnaire and examination measures for level 1 and 2 are in common use today within the fire service and provide both qualitative and quantitative data points to make decisive decisions in quality, specificity, and effectiveness of training. Metrics can easily be applied to gain important data from level 3 and level 4 but have inherent pitfalls depending on the measures used for this purpose.

[30] Kirkpatrick developed his four step model in 1959 and provided a simple and pragmatic model for helping practioners think about training programs. The majority of the work that followed is based upon Kirkpatrick's program. (excerpt from Tamkin, P., Yarnell, J., Kerrin, M., *Kirkpatrick and Beyond: A review of models of training performance*, IES, 2002, p. ix)

[31] Kirkpatrick and Beyond, p 3

b. Hamblin – Five Step Model

The first modification of Kirkpatrick's model was described by Hamblin (1974) in a five level model. He uses Kirkpatrick's original first three levels and expands upon the fourth level to account for changes training can have on the organization and upon the financial bottom line of the organization and the economy as a whole. He suggests that his model forms a hierarchy of training evaluation and that the levels can form an ascending value of information with the final level being ultimately the most important outcome, financial effects.[32]

Both authors have been criticized (Holton 1996) for not accounting for intervening variables that effect learning and the knowledge transfer process. Holton further states that no model can be validated without measuring outcomes and developing metrics for that purpose. Kraiger and Jung (in Quinones, 1997) argues that Kirkpatrick's model serves as a framework of how to evaluate training but does little on what to evaluate and how to link the results to the strategy of the organization. All the criticism of Kirkpatrick's model seems to be a constructive exercise that allows the model to be the foundation from which many other models have been built.

c. Organizational Elements Model (OEM) – Six steps

The Organizational Elements Model (OEM) suggested by Kaufman and Keller (1994) expands on Kirkpatrick's model to discuss societal influences upon outcomes. They argue that each level should include planning and needs assessment and an examination of desired or expected results. This would improve the quality of data measured so a more accurate analysis could result in a more useable outcome.

In 1995, with the help of Watkins[33], Kaufman and Keller described six levels:[34]

- **Level 1: Input** – similar to Kirkpatrick's reaction level but has been expanded to include the role, usefulness, appropriateness and contribution of the methods and resources used

[32] Tamkin, P., Yarnell, J., Kerrin, M., *Kirkpatrick and Beyond: A review of models of training performance,* IES, 2002, p 5
[33] Watkins helped the team of Kaufman and Keller produce the six step model.
[34] ibid, Tamkin, p 6

- **Level 2: Process –** the reaction level, but also includes an analysis of whether the intervention was implemented properly in terms of achieving its objective

- **Level 3: Micro (acquisition)** – the learning level plus it examines the individual as well as small-group mastery and competence

- **Level 4: Micro (performance)** – links closely to behavioral level and examines the utilization of skills and knowledge. This focus is on application rather than transfer of skills and knowledge

- **Level 5: Macro** – relates to the results level and examines organizational contributions and payoffs

- **Level 6: Mega** – an additional level which looks at societal outcomes

d. Indiana University and the Carousel of Development: Industrial Society Model (IS)

Two other models, the Indiana University approach (Molenda, Pershing and Reigheluth, 1996) and the Carousel of Development (Industrial Society 2000) continue to refine the four levels with differing emphasis upon specific levels. Molenda, et al, describes six 'Stratum'[35] and puts emphasis on the 'impact that changed performance will have on society'[36] while the Industrial Society, emphasizing six stages, aims to validate and evaluate training. "Andrew Forest of the Industrial Society argues that 'true evaluation needs to take place long before and after training has taken place'."[37]

e. Return on Investment (ROI) and KPMT[38] Models

The five-level Return On Investment (ROI) model (Phillips, 1994, Phillips and Holton, 1995) attempts to use the Kirkpatrick four-level model and add a link to monetary benefits measuring the relationship of training results to cost or what is the return on the investment we put into the training? This line of thought is expanded with the KPMT model (Kearns, P. and Miller, T., 1997). They argue that pre-measurements need to be in place to assess if a training program is appropriate. These measurements

[35] ibid, Tamkin, p 7
[36] ibid, Tamkin, p 7
[37] ibid, Tamkin, p 8
[38] ibid, Tamkin, p 8, KPMT is the first letter of the last and first names of the authors, Kearns, P. & Miller, T.

identify the need and the training program is developed for that need. The program can only bring added value to the organization if there is poor performance or there is a need identified to further the mission statement of the organization. In other words, an organization should not train for the sake of training, there must be an identifiable need and the use of metrics can be instrumental in the identification process.

f. CIRO – Context, Input, Reaction, Outcome Model

The CIRO (Context, Input, Reaction, Outcome Approach) model (Warr, Bird, and Rackham, 1970) restructures the four level approach and front loads the process by analyzing the context and inputs to the training program prior to assessing the reaction level.

g. Alternative Evaluation Models

This section will embrace alternate model areas, models that focus on evaluation and models using different types of measures. These models stray away from the four-level Kirkpatrick model and emphasize other critical areas.

Pulley (1994) described a model that places evaluation and metric development as being dependent upon what is good for the end user. He uses stages to accomplish his argument[39]:

1. Identify the decision-makers so as to ascertain who will be using the information and what their stake in it is.
2. Identify the information needs of the decision-makers – what do they need to know and how will it influence their decisions?
3. Systematically collect both quantitative and qualitative data.
4. Translate the data into meaningful information.
5. Involve and inform decision-makers on an on-going basis.

One of the alternative models describes four different educational evaluation processes (Stufflebeam et al., 1971).

1. Context Evaluation – which helps in planning and developing objectives
2. Input evaluation – which helps to determine the design by examining capability, resources and different strategies.
3. Process evaluation – which helps to control the operations by providing on-going feedback.

[39] ibid, Tamkin, p 15

4. Product evaluation – which helps to judge and react to the program attainments in terms of outputs and outcomes.

Two other alternative models continue to emphasize evaluation as a focal point. Newby (1992) argues that evaluations can and should be done during training and exercise events and after the event in the work environment, by using performance metrics. He also argues that the criteria to measure can be abstract influences that affect the ability to perform, such as societal, moral, political or philosophical criteria.[40] Preskill and Torres (1999) emphasize evaluation is a learning process that connects the organizational mission and strategic plans. They emphasize issues critical to the organization, promote dialogue, emphasize internal processes as well as outcomes, provide educational and training opportunities, promote collaboration, co-operation and participation, and use diversity of perspectives to develop understanding about organizational issues. They treat this process as a learning opportunity for the entire organization.

h. Models Using Different Measures: Evaluation Models

Kraiger et al. (1993) argues that evaluation measures will become clearer if we link learning outcomes and training evaluation. They distinguish between three different types of training outcomes, cognitive, skill-based, and affective. If we take these and look at them from the perspectives such as the goals of the training, the process strategies and the performance criteria, what the evaluation measures will be and their results will be clearer and more useful to the end user.

The one alternative model that has gained wide acceptance throughout the private, corporate world in the last eight years is the Balanced Scorecard model (Kaplan and Norton, 1996). They use measures across four areas of an organization: finance, customers, internal business processes, and learning and growth, with the first set of measures for finance being as important as the metrics for learning and growth.

3. Summary: Private Sector Concepts That Can Be Adapted to the Fire Service

[40] ibid, Tamkin, p 16

a. Kirkpatrick's model describes four levels of evaluation that should be undertaken. The fire service measures two of them routinely, the reaction level and the learning level. A large portion of our training regiment is classroom, instructor based and evaluation questionnaires evaluating the course material and level of instruction are easy and commonplace along with post-testing. Neither behavioral change nor results of the training are officially evaluated on a professional level. But, through the use of exercises, field evaluation of skill levels of individuals, units and multi-unit interaction can be observed. Also, Newby (1992) points out, that evaluations should continue after the training to ensure the effectiveness of the training. This is one concept that needs to be adapted widely by the fire service if our level of preparedness to respond to terrorist incidents that are few and far between is to be optimized.

b. The private sector reinforces the military concept of evaluation and assessment to bring the fire service to an anticipated level of preparedness. Both the IS Carousel and the KPMT models point out the necessity to identify the need through measures that point out deficiencies and gaps. This will propel us into the learning process where constant evaluation and measures need to be used to benchmark our progress. We must also take into account the individual factors that will influence our learning, positively or negatively, and exploit them. This will result in outcomes that need to be reinforced by the organization to create behavioral change. Each of these levels can be measured and benchmarked to keep the process moving forward and to provide feedback to improve process. With this accomplished, real, positive organizational change can be measured and should be measured.

c. Pulley (1994) suggests that we should identify the stakeholders and ascertain their needs. We should not use metrics for the sake of measuring: they should be purposeful, exact and sensible, for the end-user. Measures should also be simple and translate easily into meaningful data and should be both qualitative and quantitative in nature. The metrics should paint a picture of the training and the preparedness of the unit in the eyes of the commander so that he could use it to assess their true preparedness capability. .

THIS PAGE INTENTIONALLY LEFT BLANK

IV: DEVELOPMENT OF METRICS

A. INTRODUCTION

Before determining which performance metrics that should be incorporated into the FDNY exercise system, we need to explore the criteria by which such decisions should be made. The following chapter examines the specific objectives that need to be accomplished by metrics, and applies them to key issues in the FDNY exercise system, with reference to the shortfalls and best practices that were identified by the preceding two chapters.

B. CAPABILITY, PREPAREDNESS AND READINESS

Capability, Preparedness and Readiness are three terms that are often used together but are not interchangeable. These are three very important concepts for the Fire Service when it comes to the application and analysis of the metrics used during training and exercises.

Capability refers to the ability of a member or unit to effectively complete a task. The Fire Service has many units with differing capabilities. An engine company has the capability to stretch hose lines and pump water thus extinguishing the flames. A truck company has the ability to vent, enter and search (VES) thus venting the fire building, causing conditions to improve to increase survivability of rescuers and victims, forcing entry to a building to allow the stretching of the line and searching for the origin of the fire and rescue of victims. The units, operating together as a team, have the capability to extinguish the fire. Thus, it is important for the Incident Commander to know the capability of the responding companies to assign tasks.

Preparedness is the state of full readiness.[41] The Fire Service is being prepared to increase its capability to operate at terroristic events through training, equipment familiarization and exercise evaluation. It is this preparedness that the Fire Service needs to analyze through the use of metrics in order to have an idea of tasks and missions it will be able to support, the budgets it will need and the manpower it will take to complete its sphere of operations.

[41] Encarta online dictionary, www.encarta.msn.com/encnet/features/dictionary, May 2, 2004

Readiness is the state of being prepared for something.[42] In order to be ready to attempt a mission, you need to be prepared to handle the outcome of your actions. If you are at heightened readiness, then you are prepared to handle many outcomes. Readiness presupposes some kind of preparedness. All three of these concepts are measurable through performance. The measurements can be quantified or can simply be observed by qualified experts.

C. MEASURING PERFORMANCE

Performance measure is the single most important metric to gather in that it measures inherent capability. Capability is at the heart of preparedness. A commander must know how his command performs in order to assess their capability. This capability can determine their preparedness and readiness so that their future performance can be predicted. In this manner, the commander is using metrics to measure the past performance so that he can measure capability to assess preparedness and readiness at the present time so that he can predict future performance.

"…you have to know where your going and where you've been to plan how to get there…" **-Brown, et, al. (1996)**

Designing the proper measuring system for an organization and selecting the measures to use will involve answering five different types of questions.

1. Why Measure?

You can't improve what you can't (or don't) measure - Jerry Harbour

The development and use of measures for the specific objectives and goals of the course or exercise is tantamount to producing a worthwhile and successful training program. The following are reasons why better measures must be developed by FDNY:

- You cannot improve if you do not measure.
- Identify deficiencies and strengths in training protocols.
- Improve training protocols and plan for future training needs.

[42] Cambridge Dictionary Online, http://dictionary.cambridge.org, Mat 2, 2004

- Measure competency, performance and proficiency for mission essential tasks.
- Use as a basis to institute real Institutional Change.
- Integral and essential part of the training cycle.
- Provide superiors with a measurement on the success of their programs. Legitimize the training.
- Comparison to other exercises.
- Change the old way of training in the fire service.

2. Who Do You Measure?

What gets measured, gets done. - Frost, Robert

a. The Individual Participants

The main focus of small exercises or training scenarios is the individual participants. These are usually end of course training and evaluation exercises run on specific scenarios by the instructors. As the exercise grows in scale the emphasis the focus should shift to the group or the team. Here we can make a distinction between a group, which is an association of individuals whose functions are interchangeable and are brought together for a specific purpose and disbanded when the task is achieved.[43] A team[44] is different. It is two or more people with different tasks who work together adaptively to achieve specified and shared goals. The central feature is coordination. The team's functional requirements always include simultaneity[45], sequencing[46] or both. Teams usually have a history and a future and team performance models agree that there are three levels of analysis for teams: individual, team, and organizational.[47]

The question of who do you measure during an exercise depends upon what your objectives are as it relates to your overarching strategy. The fire service would need to measure several different groups depending on the goals and objectives of the exercise.

[43] Brannick, Michael T.
[44] Brannick, Michael T.
[45] Simultaneity means the team members must do something at the same time such as play notes at the same time
[46] Sequencing means that the output of one team member's task or tasks is the input to another team member's task.
[47] Brannick, Michael T.

The individual participant should be measured if competency and proficiency are important for the exercise. The middle level operational supervisor and the command chiefs would be measured if command and control issues need to be developed or improved. The interactions of agencies or jurisdictions are very important for the outcome of the exercise. The measurement of the group and team dynamic should be assessed because of the influence this dynamic has on the success of an exercise.

b. Groups and Teams

It is important to distinguish between a team and a group. The fire service operates in teams of two to six members, all performing a different function in order to complete an objective. The whole team performs as one. Metrics that measure how a team operates or how the individual operates within the team would be valuable to predict their performance.

Groups operate in pairings of two or more or groups of groups that all perform the same function to complete a task. Law enforcement and emergency medical personnel generally operate as groups of teams that perform a common task to complete an objective. The normal pairing is two individuals that form a team and then two or more of those teams form a group. Metrics that measure the interaction of that group as they relate to performing an overall objective would be important to analyze when placing these groups together in the real operational world. If they perform well during an exercise that is well planned and executed, then it can be assumed that their performance would continue under operational conditions.

c. Supervisors and Incident Commanders

Supervisors and Incident commanders are also a target of exercise metrics. Multi-agency operations and drills are interactions of differing groups of responders operating together to achieve an outcome. Exercises not only measure individual performance of operational teams, they also measure the actions of supervisors. Were the supervisors able to establish effective command and control over the operation? Did the commanders instill trust and confidence in the operational members? Measures would be very valuable to analyze if they captured the correct metrics from below, horizontally or

from above. Was incident command procedures implemented and if so were they effectively activated?

3. What Do You Measure?

> Not everything that can be counted counts, and not everything that counts can be counted. - Albert Einstein[48]

The Army calls the ability of individual operators to perform operations that are vital for the completion of their objectives or team objectives as 'mission essential tasks'. The fire service refers to this ability as 'competencies'. The ability to read and interpret readings on a meter, or the ability to operate a hand held assay, would definitely have an effect on the outcome on a situation. These would be measured at smaller training exercises or in training drills but are a main piece of the puzzle for overall performance.

> Don't measure what you can't or won't use. – Harbour, Jerry L.

A critical enabler in achieving preparedness levels and the desired level of performance from emergency responders is the ability to measure performance. In order to begin the measuring process, you must know where you are at the present time. You must establish a baseline, a beginning, a starting point. Once you know where you are, you can establish where you want to go. The direction you want to take the emergency responders along their line to preparedness. Once you have established these two points you can determine the delta or the gap between baseline and goal. Develop and implement a process that is consistent with the strategy of your organization that will get you to your goal and close the gap. It is often helpful to establish benchmarks of achievement to map out the path or re-measure periodically using the same method you used to establish the baseline.

The use of performance metrics during this process can be simple measures that establish the baselines like the measurement of time factors, such as how long it takes to

[48] Quote by Albert Einstein taken from a NPS thesis by Leighty, James E., December 2001, p. 13.

get responders onto the scene of an incident. Alternatively, the measure can be numerical (e.g., how many responders responded to the incident?) If you take a baseline and at any particular incident you find that it takes six minutes to get an ambulance to the scene of a particular incident then you have identified a gap in response. You set a goal of three minutes for a maximum response and then set about producing a process to achieve that goal. An exercise can be developed to test a realistic response time and measured. An ambulance is set at point A and it needs to respond to point B. If you take into account all the variables and control them so that they are minimized in the exercise and you achieve the three minute goal, then you know it is realistic and you strive to reproduce that in the real operational atmosphere. If you find that you can not match the goal in this exercise, then you need to identify the factors that are influencing this operation and change it by reducing its influence or creating new factors that have less of an influence. That is creating real institutional change on a simplified level.

What we are actually doing is setting a process for the transformation and blending of a set of inputs (emergency responders with their skill sets, including control of the incident, mitigation of the hazard, rescue of the victims) into a more valuable set of outputs (protection of life and property). This process can be further subdivided into a series of interrelated activities (instrument usage for identification of CBRNE, over-packing of dangerous materials, forcible entry and placing water on a fire) that can be further subdivided into individual process steps (turning a meter on, placing a hose line into operation, forcing a set of locks on a door, neutralizing a corrosive) that can all be measured and placed on a scale to analyze relative performance. Examples of certain measurable inputs and outputs are numbers of operational people put to work in a period of time for an operation. The number of victims transported and whether they are ambulatory or non-ambulatory. How many victims were triaged at the scene and how many medical workers were needed to accomplish this effort?

The key component to collecting these performance measures is to identify those measures that will actually help achieve the desired goals and objectives and then deliver those metrics to the right people at the right time (the incident commander, the chief of training, the head of the organization, your immediate supervisor) so that the measures

can be analyzed and interpreted so an action can be taken to allow the process to continue or a deficiency corrected.

There are many different uses for metrics and performance measures ranging from determining present performance levels to predicting future ones to carefully controlling an existing operation. They are used to determine the following[49]:

- **Baseline -** so that we know where we are beginning or where we are at the present moment.
- **Trending -** shows use the direction overtime so we can monitor the operation and change procedures if the process is deflecting from the chosen path.
- **Control –** to determine if we are staying within a predetermined boundary within our process.
- **Diagnostic –** metrics and performance measures can help identify problem areas in the process. Areas that are more likely to cause a breakdown of the system or delay the outputs
- **Planning –** planning performance measures answers the question, "given certain information and past performance levels, what can I predict and plan for in the future". This is the most important type of measure and analysis for the emergency responder community. This type of performance measure will drive the planning and training cycle and thus the preparedness level. This measure needs to get to the right people at the right time, either already analyzed with constructive recommendations or left for them to analyze. These will be the measures that can be the difference between a successful operation or chaos and injuries.

4. When Do You Measure?

Measurements have a tendency to measure the past – that is, the things that have happened. In order for a measure to be of value to the organization, it must transcend the dimension of time and there should be a set of measure for each location in time, the past the present and the future. Some measurement systems may very well be able to measure

[49] Harbour, Jerry L., PhD, The Basics of Performance Measurement, Productivity Press, 1997. The types of measures will be discussed in context in the next chapter.

in all three locations in time. This will give us the ability to know where we were, where we are at, and attempt to predict where we will be in the future, both short and long term.

If the challenge is to have a Squad[50] company ready to meet a certain challenge of protecting an special event that will take place at a future date, then we need to train them to take on the responsibility, exercise them to make sure they will be able to successfully complete the mission, and continue to challenge them so we can predict that they will be continually adapting to a dynamic climate. Some of the metrics we will use prior to the exercise, during the exercise and following the exercise may look like this:

- **Pre-exercise** – examine the company training records; examine the competency of all the members, and the dates of the completion of these competencies. This will give us a baseline on where we are before entering into the performance evaluation process. If we find deficiencies prior to the exercise, then we will be able to correct it with training regiments. In order to get the information, we may just use a simple evaluation form asking each member where they think they are in relation to where we need for them to be. This can also be given to his training officer in his unit to determine his competency on a specific procedure or multiple operations.
- **Present** – While the unit is running through an exercise or an activity, observers/evaluators can be used with similar check sheet or logs to determine the competency of the unit. The important thing here is to determine the metrics you need to measure to give you a clear picture of the units ability to be prepared for this special event.
- **Post** – You can use the same metrics or other metrics (have the company fill out evaluations after the exercise to rate their performance, have an after action critique) from the exercise to determine the direction in which to proceed to accomplish your goal of having the company ready for the mission.

[50] A Squad Company in the FDNY is a multi-faceted company that is trained in all aspects of the fire service, it is a truck company, an engine, CFRD trained, Hazmat Technician unit, a light rescue company, a decon engine, and a jack of all trades.

5. How Do You Measure?

Metrics are used to measure the knowledge transfer to an individual, between team members, or within an organization. They are also used to drive performance of individuals, teams and organizations. They can take various forms.

Academic measures can be a quantitative test with a measurable outcome (example: a 100 question test that is scored on a percentage right basis). Such measures can also be an evaluation process – questionnaires that evaluate the training program, the individual's knowledge base, the instructor or leadership of a group or the knowledge that was transferred. These forms of metrics bring the members and the leadership into the process and give them a stake in the success of the process. These measures are generally qualitative but can easily be transformed into a quantitative scoring methodology.

Professional measures are usually qualitative in nature and does not affix a ranking to a member. They can be a check-off style – used in a pass/fail type situation where a member can either perform a competency or a mission essential task or they need to receive further training. Most times there is no room for mediocrity and this metric can easily be quantified by tracking the number of members correctly performing the evolution. A time frame can be added to record relative speed to complete the competency. They can be a questionnaire style- these will bring the members directly into the measuring process. Questionnaires have the most dynamic format for most measures that we use. They can ask the member to evaluate his knowledge on a subject, or his attitude about a team or the leadership of the team. These are the broadest of all metrics but can be used to receive the most data points for analysis.

Whichever metric model that you formulate, you need to make it specific to the exercise that is running and must be able to measure the outcomes that the stakeholders consider most critical. They must also be simple and easy to grasp. Remember, if your metrics are not being used, then maybe you are measuring the wrong thing.

6. Where Do You Measure?

Metrics form the connection that will drive the exercise to its final outcome. They will complete the loop between the beginning, or baseline, and the training that will be

the result of the analysis of the exercise. All phases of the exercise can be measured and analyzed.

The pre-action report contains all the information on the present capability of the members participating in the exercise program. During this pre-exercise phase, this report will form the baseline of what shape the exercise will take, an evaluational exercise, a training exercise or a planning exercise. The Exercise Design Team[51] will design an exercise that will address the given need. All pertinent training metrics become part of the evaluation process. SOP's and procedures will form the goals that the participating members must demonstrate.

During the exercise, measures are created to document whether individuals or teams meet the marks or the minimum standard. These measures are both qualitative and quantitative and are recorded and formulated into an After Action Report (AAR). Recommendations are then sent to the stakeholders (supervisors, leaders, commanders or Fire Chiefs) as actionable items that can be used to direct or redirect efforts within the organization.

D. THINGS TO REMEMBER WHEN DEVELOPING METRICS

In order to be successful in deploying metrics there are a few key ideas to remember.[52]

1. Measurement drives behavior[53] A metric has to have a specific use to a real individual or organization that is responsible for a particular outcome. If we make the metric for a specific person then he/she takes responsibility for that measurement and will make sure it gets accomplished.

2. Measure what's important to the organization, the exercise, the unit performing the evolution and the stakeholders. Be careful not to allow stakeholders to exercise too much influence over the assessment system, because they may (for political reasons) not want to highlight performance gaps. It is up to the unit members and their superiors to demand the best, expect the best, and not settle for anything less.

[51] The Exercise Design Team is described in the DHS-Office for Domestic Preparedness publication, Homeland Security Exercise and Evaluation Program, available at www.ojp.usdoj.gov/odp, March 2003.
[52] Harbour, et al. p. 65
[53] Harbour, et al. p. 67

3. **Include comparative basis as part of the overall program**. It is important build metrics that are consistent from one exercise to another in order to compare and measure the delta gap between what was known and what needs to be known. Answer the question: did we close the gap and get where we need to be?

4. **Metrics should be collected, distributed and analyzed in a timely manner** for actionable items and targeted to specific people and/or audiences. Metrics are the product of an exercise that closes the loop on the process from beginning to end so that change, if needed, can be accomplished. A series of metrics and exercises can guide the path to real institutional change if needed and the metrics are heeded.

5. **Finally, metrics need to be easily understandable, and their meaning needs to be quickly and easily grasped and understood.**[54] Time is valuable and information is overwhelming. The analysis needs to be boiled down and refined before it hits the target audience. Be careful not to "over boil;" that's when you run the risk of biased reporting. It only needs to be tender, with all appropriate information from the analysis of the data provided by the metrics.

[54] Harbour et al.

THIS PAGE INTENTIONALLY LEFT BLANK

V: THE PROPOSED METRIC SYSTEM

A. INTRODUCTION

To build an overall proposal for consideration by the leadership of FDNY, this chapter specifies the specific metrics system developments that should be pursued by the organizations that need to help the leadership restructure the exercise system: the Exercise Design Team (EDT) and the Executive Committee for Preparedness Evaluation (ECPE).

B. EDT AND ECPE FUNCTIONS

The Exercise Design Team, as it is described in the Department of Homeland Security's HSEEP Program (Homeland Security Exercise and Evaluation plan)[55], is essential to be the controlling entity to bridge the gap between training and preparedness. The Team designs the exercise to evaluate the training and the planning process. They become the independent assessor of the process, and the reports and analysis of the metrics that they generate, prior to, during and post exercise are presented to the leadership and the stakeholders of the organization. This analysis forms the basis for the assessors to predict the preparedness of the organization to respond and operate at a terrorist event.

The Executive Committee for Preparedness Evaluation is comprised of an appropriate group of qualified personnel who are knowledgeable in training, operations, tactics, exercises, and the evaluation process should act as the main assessors of the capability and preparedness of the organization. Examples of these people would be high ranking officials and chief officers or any other group of qualified officers that will consistently have access to the reports and act upon their results. They would act as the Executive Committee for Preparedness Evaluation to ensure that any gap that is identified through this process is acted upon. Whatever form this committee takes for the individual municipality, it should, as a basic criteria, be comprised of the most knowledgeable and qualified members of the organization and be headed by a staff representative that has the

[55] Homeland Security Exercise and Evaluation Program, Volume I: Overview and Doctrine, U.S. Department of Homeland Security, Office for Domestic Preparedness, March 2003, p. 23

power and ability to hold entities accountable and to make the recommendations happen that will close the gap between training and preparedness.

The Exercise Design Team should be independent of Training and the Executive committee, but should issue its reports and analysis to both entities. This would ensure that exercises are designed to test and evaluate the training and planning process and limit biases that may influence the true analysis process. Any weaknesses or strengths will be identified and recorded so that the truest analysis could be considered.

The rationale for the creation of this team is the reduction of biases that could encroach upon the system. The military charges the commander with the responsibility of assessing the preparedness and readiness of the command. This is done in an insulated environment that limits opportunities for outside influence. By using multiple experts to balance the evaluation process the EDT can accomplish the same objective.

C. EXERCISE DESIGN CRITERIA FOR USE BY THE EDT AND ECPE

The Exercise Design Team should keep in mind a number of suggestions to ensure success. First, the exercises should be focused on definitive goals and objectives. The measures that will be used should be simple, measurable and specific. More is not better.

Second, the exercises should be kept simple enough to control and measure but not so simple that the participants will be insufficiently challenged. It is important that if you have members participating, they should be challenged by the tasks for which they were trained. The exercises should not intentionally set them up for failure but they should not be guaranteed success without adequate performance. There are lessons to be learned in failures and near failures as well as successes. The exercise should be used to evaluate but also reinforce the transfer of knowledge through the learning process.

Third, the design of the scenario should be specific and based in reality. The threat level of a particular event should be well researched and topical.[56] The scenario that is used should be thoroughly thought through and mapped out as to all the possible outcomes. This is an important part of the design of the exercise so that the proper

[56] A scenario needs to use agents or chemicals as they would react if released at the given time, temperature and conditions that are prevailing at the time of the exercise and not build in falsehoods that would mislead the learning process of the participants.

metrics will be applied for analysis. "Red and Blue Teaming[57]" can be an effective method of recording all possible outcomes.

Fourth, the measures that are applied should be built into each exercise and can be specific for that exercise. However, some of them should also be adaptable for a range of different exercises so that a common measure can be compared. The ideal situation is to have the flexibility to "crosswalk" through different exercises and compare the results to draw out hidden gaps in training. There are several different skill sets and knowledge bases that occur within the fire service that need to seamlessly blend together, so that fires service components can operate as a team to complete a mission. We need to measure that ability to operate as a total, interdependent force and not as independent entities that have the potential to work against each other.

Fifth, consequence management has become a multi-jurisdictional mission. Federal, State, and local assets will be drawn upon to control the incident under one Incident Management System. The main emphasis of major, multi-agency exercises should be the proper evolution of Incident Management tasking, and the expansion or de-escalation of an Incident Management protocol. This is a key element to the management of any incident.

D. TYPES OF EXERCISES TO USE WHEN DESIGNING

The HSEEP describes seven different types of exercises that an organization can use to assess their training and evaluation process as it relates to their assessment of their capabilities and preparedness to respond to a terrorist or all-hazards response event. They are:

1. Seminars – used to orient, educate and update participants on specific plans, procedures and practices within an organization. It is generally a low stress environment comprising of discussions, presentations and general morale building for an organization.

2. Workshops – used to develop a specific product, a training course, exercise scenario or program. They are generally attended by specific, skilled

[57] Red and Blue teaming is a technique the military uses to test offensive and defensive strategies. A red team will try to foil the defensive perimeters and the blue team will plan for the defense of the perimeter. Both plans are compared and analyzed to improve the performance of both offensive and defensive entities.

representatives (subject matter experts or other professionals) that will bring a skill to the table in creating the product.

3. Tabletop Exercise – usually attended by senior staff to discuss and workout key issues on a top level platform. Tabletop exercises can test policies and procedures or the planning efforts of organizations with the goal of improving these products in a controlled setting and identifying possible gaps in training, operations or planning. Tabletops are a useful tool for mid-level commanders and can also be adapted to lower operatives as a means of developing leaders.

4. Games- this is a simulation of operations that can be performed on a computer based platform involving one or more individuals or teams. Though relatively new to the fire service, the true potential of this type of exercise is just emerging.

5. Drills – these are coordinated, supervised activities that tests the competency of specific evolutions or skills, for single members or teams of members. Drills are generally focused, measured against established standards, capable of having instant feedback, usually performed in a realistic environment, and conducted in a way so that performance can be isolated and easily measured.

6. Functional Exercises – designed to test and evaluate individual or team capabilities in the context of a realistic environment with multiple tasking and interdependent operations. They are used to evaluate performance, reinforce established policies and procedures, test planning efforts, reinforce training objectives and examine inter-team, inter-agency, or inter-jurisdictional relationships.

7. Full-Scale Exercises – large scale exercises that bring all parts of the training, planning, operational, prevention and response process to a final evaluation setting. These exercises are generally used to assess the organizational ability of players in an agency, inter-agency and/or inter-jurisdictional platform. This exercise tests both operational as well as non-operational aspects of a community's preparedness capability.

Three types of the functional exercises also exist:

a. Training Exercise

This exercise is in the context of a training course for the purposes of learning a specific skill set. The participants will be told what they are learning. Then, they will then be taught what they need to learn, they will apply it in a hands-on exercise,

and then they will review what they learned. This is the repetition model in learning. They will be measured at each step and observed as to whether they did the evolution or competency correctly.

b. Evaluation Exercise

This exercise is for the explicit purpose of evaluating performance of an individual, team, unit or organization. They can be a "no notice" exercise where they participants will be brought to an exercise location and given a problem to solve without any warning. The exercise is fully measured and graded for a specific skill or skills. They can also be a full notice exercise where they individual, team, unit or organization is put through an exercise that they had full knowledge as to time, place, conditions but did not have a complete knowledge of the scenario that would be presented to them. This type of exercise is valuable in the evaluation of the whole process of training and the application of transferred knowledge.

c. Testing Exercise

This type of exercise is a full notice exercise, usually on a volunteer basis to test a plan, procedure, doctrine, or strategy. The participants are fully briefed before hand and the exercised is fully evaluated, observed and measured, not on participant performance but how the plan or procedure affects participant performance.

E. MEASURES TO USE IN THE EXERCISES

The metrics that can be used and embedded as part of the exercise should be simple, useful and focused on the goals and objectives of the exercise (which should mirror the strategic plan of the organization). They should be positioned to account for conditions pre-exercise, during the exercise and post exercise. This will ensure that the past, present and future capabilities are being measured and are accounted for in the analysis of the data. Each set of metrics should be specific for the exercise being run, should carry a common theme throughout the cycle of exercises that reflects the benchmarks of the strategic planning process.

1. Measures for Pre-exercise

During the pre-exercise phase of the planning, written examination results or final course assessments from prior training courses should be made available to the training

officer of the unit the member is assigned so that a training profile can be developed. Every member should have a training portfolio with the dates, course, instructors and final marks recorded along with evaluations from the instructors. If the exercise is going to be evaluating evolutions and competencies that are not normally observed by the unit, then a training portfolio should be sent to the unit well in advance. This should be followed up by individual questionnaires to the members of the unit that will be participating, which will provide data to quantify their self-assessment of their training. This will be repeated in the post-exercise phase to compare results.

This gives the evaluator of the exercise a baseline framework to work from when objectively evaluating performance. The written examination, evaluation of training records and the individual questionnaires will give the assessor a feel for the pre exercise state of mind that the participants are in when comparing to the end result.

2. Measures for Use During the Exercise

The evaluators and observers should be prepared with an evaluation form that can be used in a check-off or format or permit the objective recording of operations during the exercise. Quantitative measurement of specific results can also be used as a measure of time, numbers, counts of victims, rescuers, or assets used or staged.

These types of measures, both, objective recording by evaluator/observers or the actual counting of participants, should give the assessor a clear picture of what was happening during the exercise to cause a specific chain of events. All of these factors should be present in the final report of the analysis of these data points, so that the assessors can make an educated measurement of the participant's capability and preparedness.

3. Measures for the Post Exercise

Multiple evaluations might be used in the post-exercise phase. A repeat of the self assessment questionnaire and a self evaluation by the participants would enable them to report back on where they were, and what they did or didn't learn. A peer evaluation questionnaire could examine how the team performed during the exercise: what went right, what went wrong and what they can do better in the future. That sort of assessment can also aid efforts to improve the exercise scenario. A leadership evaluation allows the participants to rate their leaders as to confidence in their abilities. The leader evaluating

the unit can produce topical discussion points that can improve performance for both the leadership and the unit. A combination of these assessments can give the evaluator a sense of the participant's perception of the exercise, a sense of what went on physically in their actions, and (more important) their thinking process during the exercise. The thinking process is the part of the knowledge gained aspect of the learning process that is often overlooked. The exchange of these experiences deepens perpetuates learning. .

These measures, together with the evaluator's reaction to the exercise, become very valuable to the assessor because they sum up the evaluations and place a useful "control" for variations between groups, evaluator/observers and the participants. There are an enormous amount of data points that can be plotted from these questionnaires or "reactionaries"[58]. Submitting all these measures in an easy to read report, which compares the statistics in graphs and flow charts, can help the Executive Committee for Preparedness assess the capability and preparedness of the uniformed force.

F. ASSESSING PREPAREDNESS: DRAWING ON LEADERSHIP WISDOM

The fire service has traditionally blended experience and knowledge of operations with the ability to perform the decision making process in a high stress environment. Gary Klein in "Sources of Power," (1998), studied the ability of fire commanders' to quickly and effectively make the correct decision while being under the high stress of an operational fire or emergency. Klein describes this as the Recognition-Primed Decision Model. Klein describes their ability to fuse two processes 1) the size-up of a situation, and 2) their recognition of the proper action to take by imagining the outcome of various options. Fire commanders accomplish this fusion in seconds by pulling in their career-long experiences and accumulated knowledge.

This process is what will make our ability to measure their performance and the ability to transfer their knowledge to less experienced and future chiefs all the more difficult. But this same process will allow us to exploit the ability of senior experienced fire chiefs to use their experience, their powers of observation and our new metrics to

57 A phrase coined by Industrial Society, 2000, in the report by the IES, Kirkpatrick and beyond, p31.

effectively assess the capability and preparedness of their individuals, their units and their departments to respond successfully to terrorism events.

To bring this leadership wisdom to bear in the performance evaluation process, the Fire Service should mirror the military: we should train one level down and evaluate two levels up. The fire service should mirror this philosophy when assessing preparedness. Firefighters should be trained by their company officers and their training should be evaluated by battalion and deputy chiefs. Thus, the Exercise Design Team should be comprised of experienced and knowledgeable officers and chiefs who should submit their analyzed reports to a committee of experienced and knowledgeable chiefs to assess their capability and preparedness level. To aid them in this endeavor I propose the following system for assessment.

G. EVALUATION SYSTEM FOR THE FIRE SERVICE

1. Purpose: To provide a logical, simplified system for commanders and the Executive Committee for Preparedness to assess in concert with exercises, the capability and preparedness of their units to respond to terrorist and/or all hazard events.

2. Basic Assumptions: All members in the fire organization are trained to the Hazardous Materials Operations level as per OSHA 1910.120 and the department adheres to NFPA 472. The skill sets for all hazard response for operational level are complied with and the hazmat team meets the standards of Hazmat Technician as per OSHA 1910.120.

3. Tools: The Executive Committee for Preparedness (ECP) shall be provided with the following reports from the company commander of the participating units. Reports and evaluations on the capability of their members as assessed by the training officer of the unit, the individual training portfolios of all the members, the manpower in unit, fully manned, percentage of non-assigned members working tours in unit/month, percentage of time/month a covering officer is working in unit, and reports and evaluations on participants in exercises as representatives of their individual units.

The ECP shall be provided with the following from the evaluators and observers of the exercise, the performance evaluation of individual members and the performance evaluations of the members who formed teams to complete tasks.

The ECP shall be provided with the following from the command staff of the exercise, the performance evaluations of the unit in completing its tasking, the ability of the unit and its individuals to follow orders and complete their tasking, the ability of the unit to coordinate and work with other units, and the ability of the unit to coordinate and work with other agencies.

Also, the ECP shall be provided with a quantitative report from the evaluators on actual numbers of measurements used as compared to the total available to use. (example: In the designed scenario, there were 100 casualties that needed to be decontaminated but only 90 were actually decontaminated. That leaves 10 victims that were unaccounted for in the operation that could be spread the contamination past the operational borders. Of course the units do not know how many victims are in the original total. This measure points to a gap in operations which needs to be addressed).

Finally, the experience and observations of the Executive Committee for preparedness shall be objectively applied to evaluate the performance of the units, the Incident Command and the overall operations and Command and Control coordination of the participants.

H. STEPS TO FOLLOW WHEN ASSESSING THE RESULTS OF AN EXERCISE

Step 1: The Exercise Design Team (EDT), working independent of Training and Operations, most likely reporting to the Planning section Chief, will establish the cycle of exercises using tabletops and functional exercises as their main performance evaluation platforms. The EDT will use the standard exercise development methodology as per the National Fire Academy and the Department of Homeland Security's HSEEP. They will develop and embed appropriate performance measurements pre, during and post exercise.

Step 2: The Executive Committee for Preparedness (ECP) will meet and become familiar with the exercise, its goals and objectives, and its support of the strategic planning process of the organization.

Step 3: The EDT will, upon completion of the exercise, coordinate and submit to the ECP a complete report of all measures, pre, during and post exercise along with the completed After Action Report (AAR) for the exercise.

Step 4: The ECP shall use the reports, evaluations, knowledge of the units and their training capacity and their observational experience in observing the exercise to establish the unit's capability and preparedness level for each of the evaluation areas used in the exercise. The ECP shall also make recommendations and file a full report to the Head of Training and the Chief of Department on the outcome of the exercise and any training or operational gaps that they have identified. The stature of the ECP shall be of such nature that their recommendations shall be held in the highest esteem and implemented.

Step 5: The ECP shall place a weighted measure upon all the evaluations similar to the following:

1. Evaluation Performance Measure Index (EPMI)

The Evaluation Performance Measure is an over arching total that is used to determine the relative performance of an entity being evaluated. The Index is heavily weighted with performance measures that are used during exercises but also evaluate factors that will influence the performance.

Evaluation Performance Measure Index (EPMI)	**Points**
Manpower	5
Training	5
Equipment & equipment knowledge base	5
Leadership accountability	5
Pre-exercise evaluation	5
Exercise performance measures	40
Post exercise reactionaries	5
After Action Report	30
Total points possible	**100 points**

***Note: Each individual organization must establish the priority and point value of each of the measures and add or subtract measures in accordance to the strategy of the individual organizations.**

Once the weighted scores for each measure is deemed appropriate by the ECP, the point totals are summed and a relative score can be assigned to each performing entity and skill set. This is just a relative assignment of points used to compare performance between units, agencies and exercises. The weighted values and evaluation points will change as deemed appropriate by the ECP.
Once the total score is assigned, the ECP can group the units in a level as depicted in the following chart or plot them on a performance line.

2. Relative Preparedness Index (RPI)

Relative Preparedness Index (RPI) Levels:

Level 1 – 90-100 points	**Fully outfitted and trained unit ready to respond**
Level 2 – 80-90 points	**Fully outfitted but 10% of unit is deficient in training or untrained, unit capable of responding and operating**
Level 3 – 65-80 points	**Unit is lacking some equipment or 20% of the unit is deficient or untrained. Unit can respond and operate with caution.**
Level 4 – 50-65 points	**Unit is lacking equipment or has 50% of unit deficient or untrained, unit can respond and operate but should be teamed with a fully trained unit.**
Level 5 – 0-50 points	**Unit is ill-equipped an ill-trained, unit can respond under the extreme caution of a trained unit and the IC.**

Caution must be used when establishing and interpreting the preparedness levels. These preparedness levels are relative and relate only to specific skill areas such as fire operational response, heavy rescue response, or hazmat. A distinction must also be drawn between preparedness and readiness. Units or agencies may profess to be prepared to respond to all hazards but lack the capability on a daily basis because of manpower, trained personnel, or equipment issues, to name a few factors, to be in a state of readiness. Capability does not translate into preparedness, and preparedness does not translate into readiness. The evaluation performance measures (EPM) and the relative

preparedness index (RPI) attempt to factor in a few important measures that will allow the incident commander to best utilize the assets presented to him at an operation.

For example: a Squad company of the FDNY is a highly trained entity in numerous skill areas. These areas include but are not limited to fire operations (tenement, hi-rise, commercial, brush, industrial, private residential, multiple dwellings, vacant buildings, row-frame, Old law tenements, new law tenements)[59], confined space operations, collapse operations, water rescue and scuba diving, hi-angle rope rescue, medical emergencies, and hazmat. Some squads may be more proficient at certain areas because of their experience in response to these areas. Normal Ladder companies may be more proficient in Hi-rise fire operations with a level 1 preparedness but lack hazmat capability so they become a level 5 response units for hazmat skills. This knowledge can be of extreme importance to incident commanders while they are developing their strategies at incidents. Though all chief officers have a general idea of unit assets, those assets change from tour to tour and week to week. The measures within the relative performance measures attempt to reflect those factors that have the most impact upon preparedness and readiness.

3. Sample Calculations

Sample calculations to use when figuring out weighted points for specific measures.

Manpower in the unit calculation – weight = 10 points

Assumptions: 5members per tour working in unit

5 members/tour x 2 tours/day x 30 days/month = 300 man-tours/month

# of tours a non-assigned member is working in unit = points

[59] These are types of occupancies that are in NYC and responded to by FDNY. Each occupancy will have different tactics and procedures associated with it when battling a fire in the occupancy.

Assigned member/tour	man-tours/month	= weighted score
5	300	= 5
4 – 5	240 – 300	= 4
3 – 4	180 – 240	= 3
2 – 3	120 – 180	= 2
1 – 2	60 – 120	= 1
0 – 1	0 – 60	= 0

- The ECP shall assign the weighted score depending upon the units manpower needs. It is assumed that a unit will operate effectively as a team with a full assigned roster working but will degrade as a function of outside members working in unit.

- For this exercise the weighted score for manpower will be 10 points out of a total of 100 points available on the performance line. This weighted average will change for each exercise according to the recommendation of the ECP.

- Not all the weighted formulas are shown. Units score could vary from tour to tour and it would be a very useful measure for the Incident commander to know. All actions to either use a particular company or not would be up to the IC and his knowledge and understanding of the system.

Training weighted/unit calculation = 10 points

Assumptions: Training points are assigned for specific training areas that the units are responsible. Squads companies are responsible for the most tasks, while an engine and truck company are only responsible for fire operations, emergencies and medical responses. The ECP can select not to use this measure for certain units.

members x hours of available training = Total hrs of unit training

Specialized Rescue training available:

Rescue technician	40 hours
Confined Space	40 hours
Structural collapse	40 hours
High angle rope rescue	40 hours
Total available hours	**160 hours x # members assigned to unit = Points**

160 hours x 25 members (Squad Company) = 4000 hrs.

(4000 hrs represents a fully trained Squad Company in rescue training)

Points available:

Hours of training/unit	# of points assigned
4000 hours	5 (represents a fully capable, prepared & ready Squad)
3000 – 4000 hrs.	4 (may represent new members not fully trained)
2000 – 3000 hrs.	3 (may represent a depletion of manpower due to injuries, transfers, etc.)
1000 – 2000 hrs.	2 (may represent serious problems in manning and training)
0 – 1000 hrs.	1 (may represent a need to place this squad out of service for squad responses, this becomes a decision of the ECP))

Equipment and equipment knowledge base can be graded on whether the equipment is available and calibrated and whether there are a sufficient number of members that are competent in the use of equipment. These numbers can be gathered from company records of equipment logs and competency logs.

Leadership calculations are based upon a fully staffed, regularly assigned officers to the unit that are fully trained and competent, and that posses the full confidence of their members. The ECP shall assign a relative number according to the above criteria.

Pre-exercise evaluation is a questionnaire of all the participants that establishes their perception of their capability and performance in operations, also hi-lighting their confidence in their leadership, both unit and command. The ECP shall generate this number from the questionnaires as compared.

Exercise performance measures should be the heaviest weighted measure in that it measures the actual performance of members and units. The number is gathered from the evaluation reports and checklists of accomplishments developed for each exercise. The EDT shall gather these measures and report them to the ECP for final adjustment and analysis.

Post exercise reactionaire is similar to the pre-exercise questionnaire and is completed by all participants which is then analyzed and reported to the ECP.

After Action Report is another heavily weighted measure that takes into account the observations of the evaluators, the benchmarks achieved by the units and their operations planning and implementation.

Total Points is the output of the metric process for a specific exercise or activity that can be used to compare the performance of a particular unit, agency or of a total response. This gives the ECP a measure upon which to assess the preparedness of the organizations to respond to the exercised hazard.

Exercises are meant to be vehicles of learning and evaluation so that participants can experience new operations so that they can recall and apply them when the real incident occurs. The Fire Service must use exercises to increase competency, preparedness and readiness and not lose the opportunity to learn. They must reject the urge to compete and not be conceited. Terrorist incidents require a competent, prepared workforce that has multiple disciplines represented that can be deployed in a joint force to attack the problem head first and not allow it to become a one agency show that is a "jack of all trades and master of none". This is a recipe for failure that has been pointed out by the mediocre exercise system presently employed.

THIS PAGE INTENTIONALLY LEFT BLANK

VI: CONCLUSIONS AND RECOMMENDATIONS

A. INTRODUCTION

An important opportunity exists to improve the value of exercises in building Fire Service capabilities. This thesis has proposed a Metrics Program that can help accomplish that goal. The proposal reflects an analysis of the shortfalls that exist in the current exercise system, and builds recommendations on how to fix them by drawing on best practices from the military, private sector and – equally valuable –FDNY itself. The NEP has given us the template to help guide the restructuring of the exercise system. The military has shown us that accountability and evaluation by superiors is a hallmark that will drive the direction of training. The civilian performance models have showed us why, when and how the Fire Service can better measure performance as part of a broader system to make the Fire Service stronger.

The Exercise Design Team (EDT) and the Executive Committee for Preparedness Evaluation (ECPE) will play a key role in this process, as called for by the Homeland Security Exercise and Evaluation Program (HSEEP). This thesis specifies planning guidelines that these organizations should follow in developing and applying metrics. Exercises should be focused on goals and objectives and support the strategic plan of the organization. Exercises should be simple to control but challenging and meaningful for the participants. The scenarios should be realistic and the measures for each exercise should be specific but promote comparisons among exercises and agencies. Finally, exercises that are multi-jurisdiction or multi-agency should stress the need to follow an appropriate Incident Management System (IMS) and the deficiencies should be emphasized and the agency should be held accountable for the deficiency so that it becomes correctable. The goal is not to produce embarrassment, but rather competency and preparedness that can be objectively measures (and improved, when deficiencies are identified).

B. RECOMMENDATIONS

1. FDNY needs to develop performance metrics that provide a reliable indicator that the training and exercise system has successfully achieved knowledge continuity and

growth within the organization. Many researchers have studied the way that organizations and their members learn, the one thing these scholars all agree upon is that in order to know where we must go, we must remember where we came from and where we are at the present time so that we can assess (and improve) our performance in the future. Each step along the way requires a metric or set of metrics to guide us along the path. Moreover, every organization should have a strategic plan with goals and objectives clearly discernable, and the training that an organization develops must lead to the achievement of the organizations objectives and goals. The only way that an organization will know if it has met its goals and adhered to its strategic plan is through the metrics that it creates to measure its past, present and future performance capabilities.

2. The Fire Service has always trained its members under the most realistic conditions that it could reproduce. This is reinforced by the Army's practice of 'train as you fight' with realism in both equipment and environment. What the Fire Service lacks is creating realistic props in exercises that will challenge its members to use their experience as well as their skills to solve problems. This is especially true for homeland security-related missions (in which, I pray, few live events will provide for training experience). Doing this will present an air of urgency and danger but will allow more valuable performance evaluations. Exercises tend to be 'dumbed down' to accommodate other agencies and time restraints. FDNY exercises should be realistic and challenging to all participants. This will translate into and help facilitate accurate performance measurements.

3. Use the measures that will measure only the aspects of an exercise that are pertinent to the goals and objectives of that particular exercise. These goals and objectives should reflect the strategic plan of the organization. More measures are not better – and too many are an impediment to the effective use of exercises as a tool for performance assessment and improvement. FDNY exercises should only measure what we will use and implement.

4. Feedback from every participant should be encouraged. The feedback should be encouraged prior to the exercise, at appropriate places during the exercise and at the end of the exercise. It should also be accepted for some time after the exercise has ended, after participants have put some thought put into their experience. The military does a

good job of accepting feedback from its soldiers as well as the commanders. Feedback can and should be used by the FDNY exercise system as a mechanism to produce change.

5. In the military, the commander is responsible for assessing the capabilities of his command, supported by assessments made by his training team. The Fire service needs to adopt the practice of assessing capabilities on the unit level as well as the Commander or Chief level. The military has worked hard to isolate its training process from bias and stakeholder influence. To accomplish the same objective, the Fire Service needs to adopt a panel of members, with appropriate rank, training and experience to help provide for unbiased assessments of performance.

6. Responsibility for training is a personal as well as a command responsibility. There must be accountability in the ranks to produce a competent workforce that can adapt and respond to any operational circumstance. Measurement produces a pride that will increase performance. The Fire Service should adopt the military practice of training one level down and evaluating two levels up. This produces accountability up and down the ranks that results in a more cohesive, competent work force. The members know who they are accountable to and the commanders know what they are accountable for within their command.

7. Training standards must be produced and followed. The military has training standards for the Force that they adhere to when training. Presently there are existing standards for hazardous materials response and mitigation that need to be reinforced by all agency participants so that skill sets can be compared and evaluated. The Fire Service must support the requirement to establish standards for all training.

8. After Action Reviews (AAR) must become the standard after the completion of exercises so that good practices can be reinforced and deficiencies identified and corrected. Deficiencies should be addressed immediately. The AAR should be focused on key competencies derived from the training objectives. Emphasis should be placed on the correct procedure to be followed, rather than pronouncing judgment on success or failure. Leading questions should be asked to encourage participants to self-discover important lessons from the exercise, and the AAR should be open to everyone who participates.

9. Stakeholders in the event should be identified but not given undue influence over an exercise. Exercises have the tendency to be tailored to certain organizations to bolster their apparent performance. By adopting the approach to metrics proposed in this thesis, the risks of stakeholder interference can be minimized.

C. TOPICS FOR FUTURE RESEARCH

Improvements can always be made on evaluation processes. This thesis has outlined multiple performance evaluation models that exist in the private sector. They can be used in multiple ways to measure preparedness in the Fire Service. As the Fire Service becomes increasingly specialized, newer and better training methodologies will be explored. Metrics will be needed to measure their effectiveness in improving competency and knowledge transfer and accumulation. Similar methodologies need to be implemented for the other disciplines that respond to the consequences of a terrorist activity so that they can be coordinated and combined under the Incident Management System (including Law Enforcement, Medical and Federal response agencies).

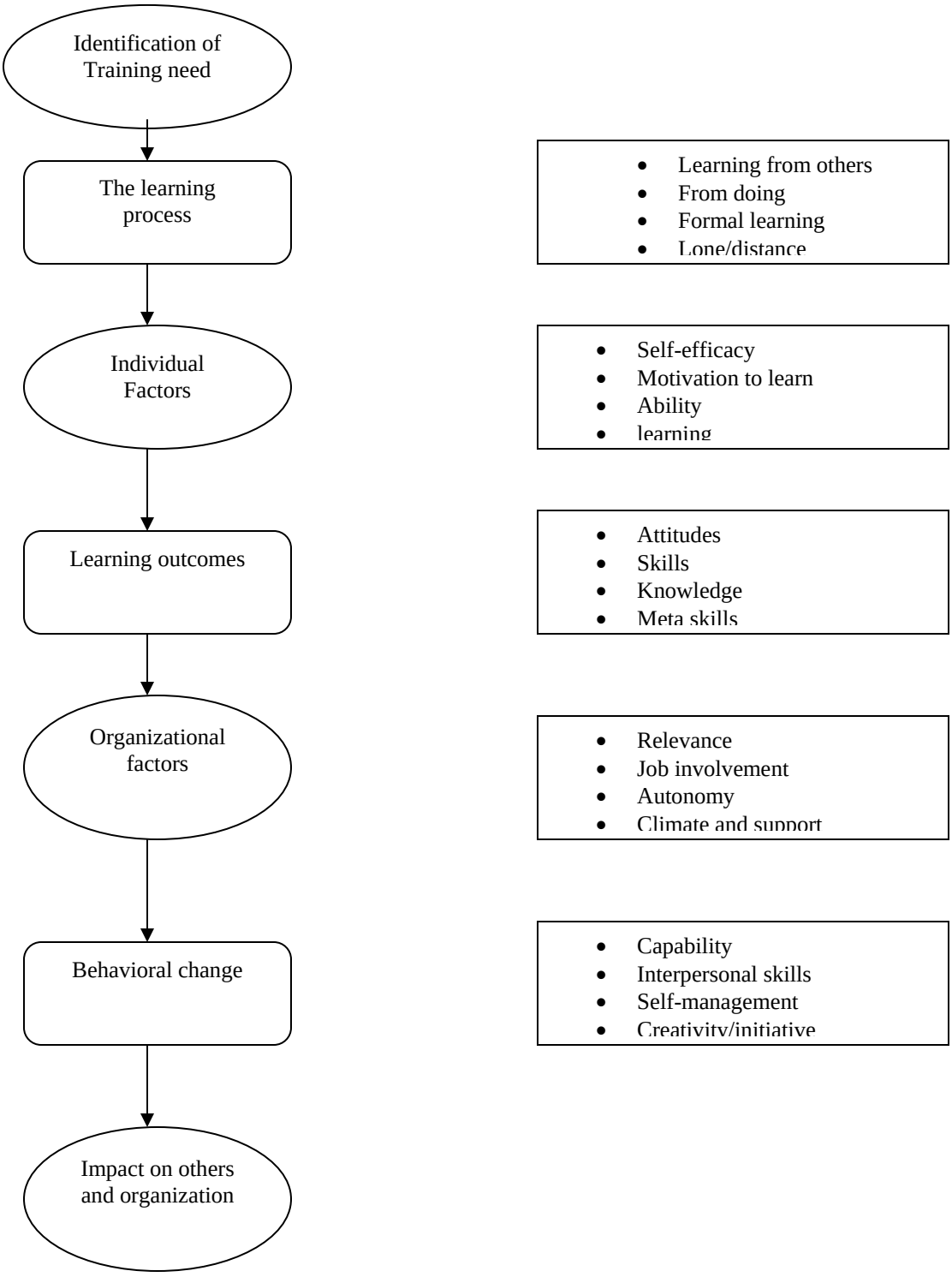
Table 2: Comparison of different models of training evaluation[60]

Kirkpatrick	Hamblin	OEM Model	University	IS Carousel	Phillips, ROI	KPMT	CIRO
			Activity Accounting	Identity Business need; define Development Objective; design Learning process		Staged process to examine the business needs; design solutions and get buy-in	Context analysis Input
Reaction	Reaction	Input Process	Reaction	Experience the Learning process	Reaction and planned action	Reaction to training and development	Reaction
Learning	Learning	Micro-acquistion	Learning	Use and Reinforce the learning	Learning	Learning	Outcome immediate
Behavior	Job behavior	Micro Performance	Transfer of Learning		Job application	Transfer to the workplace/ behavior	Outcome intermediate
Results	Organization	Macro	Business impact	Judge the Benefits to the organization	Business results	Bottom line added value	Outcome Ultimate
	Ultimate value	Mega Societal Outcomes	Social impact		Return on investment		

[60] Tamkin, P., Yarnell, J., Kerrin, M., "Kirkpatrick and Beyond: A Review of Models of Training Evaluation", IES, report # 392, 2002, p13

THIS PAGE INTENTIONALLY LEFT BLANK

Table 3: A model of evaluation[61]



[61] Tamkin, P., Yarnell, J., Kerrin, M., "Kirkpatrick and Beyond: A Review of Models of Training Evaluation", IES, report # 392, 2002, p29

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Brannick, M.T., *Team Performance Assessment and Measurement: Theory, Methods, and Applications,* Lawrence Erlbaum Associates, Mahwah, NJ, 1997

Bryson, J.M., *Strategic Planning for Public and Nonprofit Organizations: A guide to strengthening and sustaining organizational achievement",* Jossey-Bass, San Francisco, CA, 1995

Brown, G., *Keeping Score: Using the Right Metrics to Drive World-Class Performance,* Quality Resources, New York, 1996

Fire Department, City of New York, *FDNY Strategic Plan, 2004-2005,* January 1, 2004

Frost, R., *Measuring Performance: Using the new metrics to deploy strategy and improve performance,* Measurement International, Dallas, TX, 2000

Hamblin, A.C., Evaluation and Control of Training, McGraw Hill, 1974

Harbour, J.L., *The Basics of Performance Management*, Quality Resources, New York, 1997

Holton, E.F., *The flawed four-level evaluation model,* Human Resource Development Quarterly, 1996, Vol. 7(1), pp.5-21, Spring

Industrial Society, *Managing Best Practice: Training Evaluation,* no. 70, April, 2000

Interview between Lieutenant John Regan, Executive Officer of Probationary Firefighter School, New York City Fire Academy, Randall's Island, New York, and the author, February 9, 2004

Interview between Battalion Chief William Hines, Training Officer, Battalion 7, Division 1, FDNY, Lower Manhattan, and the author, May 19, 2004

Kaplan, R.S., Norton, D.P., *Focusing Your Organization on Strategy – with the Balanced Scorecard, 2nd Edition,* Harvard Business Review OnPoint Collection, Product 5933, Harvard Business School Publishing Corporation, 2004

Kaplan, R.S., Norton, D.P., *Having trouble with Your Strategy? Then Map It,* Harvard Business Review OnPoint Collection, Product 5165, Harvard Business School Publishing Corporation, 2004

Kaufman R., Keller, J.M., *Levels of Evaluation: beyond Kirkpatrick,* Human Resource Development Quarterly, Winter, 1994, Vol. 5(4)

Kaufman, R., Keller, J.M., Watkins, R., *What works and what doesn't: evaluation beyond Kirkpatrick,* Performance and Instruction, 1995, Vol. 35(2), pp. 8-12

Kirkpatrick, D.L., *Four steps to measuring training effectiveness*, Personal Administrator, 1983, Vol. 28(11), pp. 19-25

Kirkpatrick, D.L., *Great ideas revisited: revisiting Kirkpatrick's four-level model*, Training and Development, 1996, Vol. 50(1), January, pp.54-57

Klein, G., *Sources of Power: How People Make Decisions*, The MIT Press, Cambridge, MA, 1999

Kraiger, K., Jung, K., *Linking training objectives to evaluation*, in Quinones, M.A., Ehrenstein, A., (eds), *Training for a Rapidly Changing Workplace: Applications of Psychological Research*, American Psychological Association, 1997, Washington DC

Leighty, J. E., *Criteria for Evaluating United States Marine Corps Installation Strategic Management*, Master's Thesis, Naval Postgraduate School, Monterey, CA, December 2001

Meyer, M. W., *Rethinking Performance Measurement: Beyond the Balanced Scorecard, The Wharton School*, University of Pennsylvania, Cambridge University Press, 2002

Merriam-Webster On-line Dictionary, www.m-w.com/cgi-bin/dictionary, May 5, 2004

Molenda, M., Pershing, J., Reigeluth, C., *Designing instructional systems, in Craig, R., (ed), The ASTD Training and Development Handbook: a Guide to Human Resource Development* (4th edition), New York, McGraw Hill, 1996

OSHA 29CFR 1910.120: General Industry Regulations, Mangan Communications, Inc., Updated through April 1, 2003

Quinones, M., (ed.), *Training for a Rapidly Changing Workplace: Applications of Psychological Research*, American Psychological Association, Washington, DC, 1997

Tamkin, P., Yarnell,J., Kerrin, M., *Kirkpatrick and Beyond: A review of models of training evaluation*, The Institute for Employment Studies (IES), Report 392, Mantell Building, Falmer, Brighton, BN1 9RF, UK, 2002

Stockton, Paul, *Making a Difference: Facilitating Organizational Learning and Change Through The National Exercise Program*, Draft Report, A joint pre-proposal to the ODP Exercise Division by Stanford University and the Naval Postgraduate School, September 30, 2003

United States Army, *Training the Force*, Field Manual No. 7-0 (FM 25-100), Headquarters, Department of the Army, Washington, DC, 22 October 2002

United States Department of Homeland Security, Office for Domestic Preparedness, *Homeland Security Exercise and Evaluation Program*, March 2003

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor Paul Stockton
   Naval Postgraduate School
   Monterey, California

4. Professor Rudy Darkin
   Naval Postgraduate School
   Monterey, California

5. Commissioner Nicholas Scoppetta
   Fire Commissioner
   Fire Department of New York City
   Brooklyn, NY

6. Chief Frank P. Cruthers
   Chief of Department
   Fire Department of New York City
   Brooklyn, NY

7. Chief Michael Weinlein
   Assistant Chief, SOC
   Fire Department of New York City
   Brooklyn, NY

8. Chief Robert J. Ingram
   Chief in Charge, Haz Mat Operations
   NYC Fire Academy
   Randall's Island, NY

9. Chief Thomas Galvin
   Chief in Charge, Bureau of Training
   Fire Department of New York City
   Brooklyn, NY

10. Chief Nicholas Santangelo
    Chief in Charge, Fire Academy
    NYC Fire Academy
    Randall's Island, NY

THIS PAGE INTENTIONALLY LEFT BLANK